# 应用随机过程

钱敏平　龚光鲁　著

北京大学出版社
北　京

书　　名：应用随机过程
著作责任者：钱敏平　龚光鲁
责 任 编 辑：王明舟
标 准 书 号：ISBN 7-301-03894-1/O・423
出 版 者：北京大学出版社
地　　址：北京市海淀区中关村北京大学校内　100871
网　　址：http://cbs.pku.edu.cn/cbs.htm
电　　话：出版部 62752015　发行部 62754140　编辑部 62753160
电 子 信 箱：zpup@pup.pku.edu.cn
排 印 者：北京大学印刷厂
发 行 者：北京大学出版社
经 销 者：新华书店
850 毫米×1168 毫米　32 开本　12 印张　290 千字
1998 年 10 月第一版　1998 年 10 月第一次印刷
定　　价：20.00 元

# 前　言

近年来，有关随机过程的基本知识与方法，不仅为概率统计专业所必需，也为计算数学、物理、生物、化学等自然科学，以及工程、经济、管理，以至人文科学等广阔领域的研究工作所需要．为适应这种需要，各高校不仅为概率统计专业大学本科生开设应用随机过程课程，也纷纷为上述其他专业的研究生或高年级大学生开设应用随机过程课．笔者与许多这方面的任课教师讨论、交换意见，同感缺少一本适用于此类课程的教材，因而，尽管深知编写这样一本教材是十分艰难的，迫于教学的急需仍然勉为其难作一尝试．

鉴于上述写作意图与读者对象，本书着重于随机过程的概念、思想方法、计算与实例，而回避测度论水平的数学严格性．希望使只具有高等数学及初等概率论基础知识的读者就能阅读与学习其主要部分．书中努力收集各方面的应用模式，将笔者见到的文献资料中好的应用实例加工、简化为宜于教学的材料，写入正文或列入习题．特别如模拟退火——优化的概率方法、神经网络的随机模型、隐马氏模型、吉布斯抽样等新近迅速发展起来的方向是随机过程的极好应用，本书力图简要地阐明其方法与思想，对读者起一个抛砖引玉的作用．为了便于有兴趣的读者进一步学习，每章后设有附录，指出相应的尚待深入讨论或严格化的理论与实际问题及参考书与文献．

已故的我国概率统计专业的开创者许宝禄教授指出：与物理或其他领域对随机现象研究不同，概率论中随机变量的概念中暗含随机自变元 $\omega$——基本事件（或者说轨道），这个 $\omega$ 的引入至关重要，它使许多含糊不清的问题，争论不休的悖论得以澄清与明晰；使许多复杂的问题得到正确分析的起点．但是在许多问题中，

人们又并不去仔细追究这些 $\omega$ 究竟是什么，只是明确有此变元，从而得到概率空间的明确概念. 本书虽则回避测度论的严格性，但仍引入概率空间的准确定义，使得我们的研究具有概率论的韵味.

全书末尾设统一的参考文献目录，按作者的英语或汉语拼音字母次序统一排列. 本书是以北京大学概率统计系本科生的"应用随机过程"课的讲义为基础，扩充、增补与修改而成的. 本书中不少内容是笔者直接由文献改写的，错误与不妥必定不少，恳请读者批评与指正，以利修改订正.

笔者得以完成本书，首先要感谢北大概率统计系，由于他们的安排和鼓励，使得笔者有机会在教学实践中完成本书的写作和修改. 还要感谢各次听课的同学，由于他们仔细阅读原讲义而帮助我们纠正了一些不妥之处. 特别要感谢陈大岳同志，他在使用原讲义的教学实践过程中发现了不少不妥之处，并提出了许多宝贵意见. 刘朝峰、张奎和李晓满等同志对本书的作图排版打印，付出了辛勤的劳动，在此深表谢意.

# 目　录

# 第一章　概 论 与 例

## §1.1　什么是随机过程？

在概率论中，我们看到随机变量的概念是很重要的，但是在大量实际问题中，往往还需要研究随机变量是怎样随时间参数而变化的．也就是说要考虑依赖参数的一族随机变量．

将一个电话交换机在$[0,t]$($t>0$)时间区间内收到的呼唤次数记为$\xi_t$，它是一个随机变量，但随$t$变化，它又是时间$t$的函数．又例如将某城市在第$n$个单位时间的人口总数记为$x_n$($n=0,1,2,\cdots$)，那么要研究此城市人口发展变化就要将$\{x_n;n=0,1,2,\cdots\}$这一族随机变量作为整体来研究，它是一个离散参数的随机过程．又如在研究流体的湍流时，将时刻$t$流体在位置$(x,y,z)$的速度记为$\boldsymbol{u}(t,x,y,z)$($t>0,x,y,z$为实数)，它就是一个随机向量，但研究湍流时就需要对不同的时间$t$与不同位置的速度作为一个整体来研究．

**定义 1.1(随机过程)**　设$T$是一个指标集，随机变量族$\boldsymbol{x}=\{x_t;t\in T\}$称为随机过程，其中$x_t$(固定$t$)是定义于概率空间$(\Omega,\mathscr{F},P)$上，取值于集合$S$的随机变量(向量)．$S$称为该随机过程的状态空间．

在上面的定义中$T$通常可取为$Z_+$(非负整数集)，$Z$(整数集)，$R_+$(非负实数集)，$R$(实数集)及它们与一个区间$[a,b)$的交．也可以取为$Z_+^n$($n$元非负整数集)，$R_+^n$($n$元实数集)，$Z^n$及$R^n$等，当$n>1$时，相应的随机过程通常叫随机场．本书主要讨论$n=1$的情况，它与$n>1$的情况的重要区别是：$n=1$时$T$有一个天然的"全序"，它与随机过程的发展是一致的，而$n>1$时$T$中不存在这

样一个全序. 当 $T$ 是 $Z^n$ 的子集时,称相应的随机过程是离散参数的;而 $T$ 当是 $R^n$ 的子区间、矩形或超矩形则称之为连续参数的.

对应于每个固定的样本点 $\omega\in\Omega$,定义 1.1 中的 $\{x_t(\omega)\}$ 就是一个以 $T$ 为定义域取值于 $S$ 的向量值函数,我们通常称之为此随机过程的一条轨道,或一个实现.

## §1.2 随机过程的分布与坐标过程

正如从概率论的观点研究随机变量(向量)着重于研究它的分布一样,对随机过程我们关心的也是它的统计特性—分布. 但是这里我们面对的是无穷多个随机变量,表征它们的"联合分布",就不能简单地照搬有限个随机变量的做法,只把个数改为无穷如法炮制.

**例 1.1** 考虑 Bernoulli 序列 $\xi\triangleq\{\xi_n:n=0,1,2,\cdots\}$,其中 $\xi_n$ 是在相互独立地进行的一系列掷币试验中第 $n$ 次掷得的结果:$\xi_n=1$ 表示掷得正面,$\xi_n=-1$ 表示掷得反面. 如果按有限个随机变量联合分布那样,考虑概率:

$$\pi\triangleq P(\xi_0=\omega_0,\xi_1=\omega_1,\cdots,\xi_n=\omega_n,\cdots)$$

$(\omega_i=\pm1,i=0,1,\cdots)$是不行的. 事实上,由于钱币正反对称,

$$P(\xi_n=1)=P(\xi_n=-1)=\frac{1}{2},$$

于是

$$\pi=\lim_{n\to\infty}P(\xi_0=\omega_0,\xi_1=\omega_1,\cdots,\xi_n=\omega_n)=\lim_{n\to\infty}\left(\frac{1}{2}\right)^{n+1}=0.$$

这样,无论试验结果 $\omega\triangleq(\omega_0,\omega_1,\cdots,\omega_n,\cdots)$ 是什么都得到零概率,从而找不到任何有意义的统计规律来表达与 $\xi$ 有关的可观测事件的概率. 自然地,我们转而考虑在任意有限个时刻的随机变量的联合分布

$$f_{t_1,t_2,\cdots,t_m,a_1,\cdots,a_m}\triangleq P(\xi_{t_k}=a_k;k=1,2,\cdots,m)$$

$$=\left(\frac{1}{2}\right)^m \tag{1.1}$$

($m\geqslant 1, i_1, i_2, \cdots, i_m$ 为 $m$ 个不同非负整数，$a_1, a_2, \cdots, a_m$ 各为 $+1$ 或 $-1$). 这一族分布决定了 $\xi=\{\xi_n; n=0,1,2,\cdots,n,\cdots,\}$ 有关的全部可观测(有概率的)事件的概率.

**例 1.2** 考虑流体的速度 $\boldsymbol{u} \triangleq \{u(t,x,y,z)(t\geqslant 0, x, y, z$ 为实数$\}$，我们也考虑流体在有限个时刻 $t_1, t_2, \cdots, t_m$ 及有限个位置 $(x_1, y_1, z_1), (x_2, y_2, z_2), \cdots, (x_n, y_n, z_n)$ 的速度的联合分布，并用这一族分布：

$$F_{t_1,\cdots,t_m,x_1,y_1,z_1,\cdots,x_n,y_n,z_n}(v_{1,1,1}, v_{1,1,2}, \cdots, v_{3,m,n})$$
$$\triangleq P(u_l(t_i, x_j, y_j, z_j) \leqslant v_{l,i,j}, l=1,2,3; i=1,\cdots,m, j=1,\cdots,n) \tag{1.2}$$

(其中 $t_i\geqslant 0, x_j, y_j, z_j$ 为实数；$u(t,x,y,z)=(u_1(t,x,y,z), u_2(t,x,y,z), u_3(t,x,y,z))$) 来刻画 $\boldsymbol{u}$ 这个随机场的统计特性.

其实，在通常的应用问题中，人们往往是先由分析或经验，得到如上的随机过程在有限个不同参数值处的那些随机变量的联合分布，而并未事先已知存在一个概率空间 $(\Omega, \mathscr{F}, P)$ 及其上定义的随机变量族(这一点在以下我们处理具体随机过程的例子时还会再仔细分析). 事实上，正如初等概率论中把 $\Omega$ 取作一切可能结果的集合一样，我们这里对例 1.1，自然可取

$$\Omega=\{\omega=(\omega_0, \omega_1, \omega_2, \cdots, \omega_n, \cdots); \omega_i=\pm 1, i=1,2,\cdots\}.$$

对例 1.2，可取

$$\Omega=\{\omega=(\underline{v}_{t,x,y,z}; t\geqslant 0, x, y, z\in R); \underline{v}_{t,x,y,z} = (v^{(1)}_{t,x,y,z}, v^{(2)}_{t,x,y,z}, v^{(3)}_{t,x,y,z})\},$$

其中每个 $\omega$ 可看成是 $[0,+\infty)\times R^3$ 上一个三维向量值的函数. 于是，一个很基本的问题就是：已知上面那样的有限维分布族，是否一定可以设法构造 $\Omega$ 上一个概率空间 $(\Omega, \mathscr{F}, P)$ 及其上的随机变量族(随机过程)，使得它在任意有限个参数值的那些随机变量具有上述已知的有限维联合分布. 从理论上，著名的 Kolmogorov 定

理可以相当一般地对此作肯定的答复. 由于这个问题不可回避地要用到测度论的知识，我们在下面讨论中不再仔细论证，只是承认这个事实，并在概率空间已经构造好的基础上去作进一步研究.

从另一个角度看随机过程，它也可当作以函数为取值的随机变元，也就是说，对固定的现实 $\omega$，$x(\omega)$ 可以看成一个从参数集 $T$ 到状态空间的映射；$x$ 以多大的概率随机地取值于不同的函数（映射），则是由上述有限维联合分布决定的. 总之，一个随机过程

$$x = (x_t(\omega); t \in T)$$

当固定 $t \in T$ 是一个随机变量（向量）；而固定 $\omega \in \Omega$，$x_\cdot(\omega)$ 是一个 $T$ 到状态空间的函数（映射）；它的统计规律由全体有限个参量的随机变量组的联合分布唯一决定.

## §1.3 简单对称随机徘徊及其坐标过程

本节中，我们以简单对称随机徘徊为例构造随机过程，以使读者对随机过程有更清晰的认识.

一个简单随机徘徊，即一个粒子在直线上每隔单位时间独立地向左或向右以均等的机会走一步. 若考虑第 $n$ 步时粒子所在的位置，它就是一个随机过程. 但是为了将这样一个直观的物理模型纳入上面的数学框架中，就需要给出概率空间 $(\Omega, \mathscr{F}, P)$，并定义其上的随机变量族 $\{\xi_n\}$ 来表示上述随机过程.

首先考虑概率空间，按照基本事件就是一切可能的随机试验的结果这一想法，对随机徘徊的一切可能结果的集合是

$$\Omega = \{\omega = (\omega_0, \omega_1, \cdots, \omega_n, \cdots); \omega_i \text{ 是整数}\},$$

其中 $\omega_n$ 表示走了第 $n$ 步粒子所在位置，由于每次只能走一步，自然若取一步的步长为单位，$\omega_n$ 只能取整数值. 令 $\mathscr{F}$ 为包含一切由有限个时间的位置限定的事件类的全体事件

$$\mathscr{C} = \{C = \{\omega: \omega_{t_1} = l_1, \omega_{t_2} = l_2, \cdots, \omega_{t_m} = l_m\}; m \geq 1, l_i \text{ 为整数}\} \tag{1.3}$$

的最小事件体($\sigma$-代数),记为$\sigma(\mathscr{C})$. 再按Kolmogorov定理(见§1.4)决定一个$\mathscr{F}$上的概率测度$P$,使得$\mathscr{C}$中的集合与上述的物理模型符合,即当对$\forall\ k, l_k-l_{k-1}+(i_k-i_{k-1})$为偶数时

$$P(\omega;\omega_{i_1}=l_1,\omega_{i_2}=l_2,\cdots,\omega_{i_m}=l_m)=p_1\cdot p_2\cdots p_m, \tag{1.4}$$

其中

$$\begin{aligned}&\omega_0=l_0,\quad p_k=C_n^m\left(\frac{1}{2}\right)^m,\\&n=i_k-i_{k-1},\ m=\frac{(l_k-l_{k-1}+n)}{2}.\end{aligned} \tag{1.4$'$}$$

而其他情形(1.4)左边的概率为0. 这是因为从时刻$i_{k-1}$到$i_k$,其中$m$步向右,$n-m$步向左($n=i_k-i_{k-1}$),总的位移是$l_k-l_{k-1}=2m-n=m+(-(n-m))$. 这个事件的概率按二项分布应为(1.4)$'$中给出的$p_k$,而由于独立性,(1.4)中的联合分布应为各$p_k(k=1,2,\cdots,m)$的乘积. 于是,我们有了概率空间$(\Omega,\mathscr{F},P)$. 若

$$\omega=(\omega_0,\omega_1,\cdots,\omega_n,\cdots)\in\Omega,$$

令

$$\xi_n(\omega)=\omega_n\quad(n\geqslant 0), \tag{1.5}$$

我们就得到了一列随机变量$\{\xi_n(\omega);n\geqslant 0\}$,它正是我们要的随机徘徊的数学表示.

对于概率空间中满足(1.5)的随机过程叫做此概率空间的坐标过程.

上面我们得到了随机徘徊的一个坐标过程表示. 当然随机徘徊并不一定要是一个坐标过程. 例如,按例1.1中的Bernoulli序列我们也可以得到一个坐标过程:令

$$\Omega_1=\{\omega=(\omega_0,\omega_1,\cdots,\omega_n,\cdots);\omega_n=\pm 1,\forall\ n\geqslant 0\}; \tag{1.6}$$

及

$$\begin{aligned}\mathscr{C}=\{C=\{\omega;\omega_{i_k}=a_k;k=1,\cdots,m\};\\ m\geqslant 1,a_k=\pm 1(k=1,\cdots,n)\},\end{aligned}$$

$$P(C)=P\{\omega;\omega_k=a_k;k=1,\ldots,m\}=\left(\frac{1}{2}\right)^m.$$

$\mathscr{F}$ 为包含 $\mathscr{C}$ 的最小事件体. 同样由 Kolmogorov 定理，按(1.1)可唯一决定$(\Omega,\mathscr{F})$上的一个概率测度 $P$，从而得到概率空间$(\Omega,\mathscr{F},P)$. 对$\forall\ \omega\in\Omega$，令

$$\zeta_n(\omega)=\omega_n\quad(\omega=(\omega_0,\omega_1,\cdots,\omega_n,\cdots))\quad(n\geqslant 0),$$

于是随机变量族 $\zeta=\{\zeta_n;n=0,1,\cdots\}$ 就是刻画 Bernoulli 序列的数学模型的一个坐标过程.

事实上，若 $\{\zeta_n;n=0,1,\cdots\}$ 是例 1.1 中的贝努里序列，令

$$\eta_n(\omega)=\eta_0(\omega)+\sum_{k=1}^{n}\zeta_k(\omega),$$

则 $\eta=\{\eta_n;n=0,1,\cdots\}$ 也是刻画随机徘徊的一个随机过程，但它不是坐标过程.

## §1.4 附　　录

1. 一个事件体($\sigma$-代数)$\mathscr{F}$ 指某集合 $\Omega$(全体基本事件)的一个子集类，满足：

$F.1.$　$\mathscr{F}$ 含空集$\varnothing$与全集 $\Omega$.

$F.2.$　对集合的可列并封闭(若 $A_n\in\mathscr{F}(n\geqslant 1)$，则 $\bigcup\limits_{N=1}^{\infty}A_n\in\mathscr{F}$).

$F.3.$　对集合余封闭(若 $A\in\mathscr{F}$，则 $A^c\in\mathscr{F}$).

2. 包含集类 $\mathscr{C}$ 的最小事件体 $\mathscr{F}=\sigma(\mathscr{C})$指一个事件体满足条件：$\mathscr{C}\subset\mathscr{F}$，而且若有事件体 $\tilde{\mathscr{F}}\supset\mathscr{C}$，则 $\mathscr{F}\subset\tilde{\mathscr{F}}$. 显然含 $\mathscr{C}$ 的最小事件体唯一.

3. 设 $\Omega=\{\omega=\{\omega_t;t\in T\};\omega_t\in\mathscr{S}\}$，$T$ 可为实数集 $R$，非负实数集 $R^+$，整数集 $Z$，非负整数集 $Z^+$，状态空间 $S$ 为可列集. 定义柱集类 $\mathscr{C}=\{\{\omega;\omega_{t_k}=a_k,k=1,2,\cdots,n\};n\geqslant 1,a_k\in S,t_k\in T\}$.

若 $\mathscr{F}$ 是包含 $\mathscr{C}$ 的最小事件体 $\sigma(\mathscr{C})$，则 $\mathscr{F}$ 上的任一概率 $P$ 由它在 $\mathscr{C}$ 上的取值唯一确定.

4. 对 3 中的 $\Omega$ 与 $T$，当 $S$ 为实数空间，或 $d$-维实向量时，柱集类

$$\mathscr{C} = \{\{\omega; \omega_{t_k} \leqslant x_k; k = 1,2,\cdots,n\}; \\ n \geqslant 1, t_k \in T, x_k \text{ 为 } d \text{ 维实向量}\}$$

（其中向量 $\omega_{t_k} \leqslant x_k$ 表示各分量分别满足不等式），则 3 中结论仍成立.

5.（Kolmogorov 定理）　设有分布函数族

$$\{F_{t_1,t_2,\cdots,t_n}(x_1,\cdots,x_n); n \geqslant 1, x_i \in R^d, t_i \in T, i = 1,\cdots,n\}$$

（其中 $R^d$ 表示 $d$ 维实数值向量）满足相容性条件

K.1）$F_{t_1,t_2,\cdots,t_n}(x_1,x_2,\cdots,x_n) = F_{t_{m_1},t_{m_2},\cdots,t_{m_n}}(x_{m_1},x_{m_2},\cdots,x_{m_n})$ 对 $(1,2,\cdots,n)$ 的任意排列 $(m_1,\cdots,m_n)$ 成立（意思是把时间次序与自变量次序同时交换，应有相同的概率分布值）；

K.2）$F_{t_1,t_2,\cdots,t_n}(x_1,\cdots,x_n) = F_{s_1,s_2,\cdots,s_m}(y_1,\cdots,y_m)$，其中 $m < n$，$\{s_1,\cdots,s_m\}$ 是 $\{t_1,\cdots,t_n\}$ 的子集，

$$y_k = \begin{cases} x_k, & \text{当 } t_k \in \{s_1,\cdots,s_m\}; \\ +\infty, & \text{当 } t_k \overline{\in} \{s_1,\cdots,s_m\} \end{cases}$$

（意思是在 $s_1,\cdots,s_m$ 中抹去若干时刻得到 $t_1,\cdots,t_n$，则在 $t_1,\cdots,t_n$ 时刻的联合分布应与 $s_1,\cdots,s_m$ 各时刻 $(m > n)$ 的联合分布在被抹去的时刻不加限制 $(x_k = +\infty)$ 的值相同).

于是存在 $(\Omega, \mathscr{F})$ 上唯一的概率测度 $P$，在 $P$ 下对于 $\{\xi_t(\omega) \triangleq \omega_t (t \in T)\}$，$(\xi_{t_1},\cdots,\xi_{t_n})$ 的联合分布为 $F_{t_1,\cdots,t_n}$，其中

$$\Omega = \{\omega = (\omega_t; t \in T); \omega_t \in R^d\},$$

$\mathscr{F}$ 是含 $\Omega$ 的一切柱集的最小事件体（$\sigma$-代数）（柱集定义见 3 与 4).

Kolmogorov 定理告诉我们，从物理或直观模型的分析得到的有限维联合分布族，可以构造一个概率空间，及其上的坐标过程，

使其有限个时间参数的那些随机变量的联合分布与物理或直观的分析所得的一致．即这个数学模型不会引起矛盾．

6．本节的严格的理论需有测度论的基础，具体可参见：汪嘉冈：《现代概率论基础》(复旦大学出版社，1988)与钱敏平，龚光鲁：《随机过程论》(北京大学出版社，1997)．

## §1.5 小　　结

本章以简单随机徘徊为例阐明随机过程的概念，即描述一族依赖于时间参数或空间参数随机地变化的量的数学模型．也就是把随机过程看成是一个概率空间上的一族依赖时间参数的随机变量．并进而指出描述随机过程的统计特征的是所有对应有限个时间参数的有限个随机变量的联合分布族．同时还指出了从实际问题得到概率空间和随机过程的建模思想．

## 习　　题

1．试构造 Bernoulli 序列(例 1.1)的概率空间$(\Omega,\mathscr{F},P)$及坐标过程．

2．设粒子作非对称的随机徘徊：每步向右或向左随机地走一步，第$k$步向左与向右走的概率分别是$p_k$与$q_k$($p_k,q_k\geqslant 0,p_k+q_k=1$)．试构造概率空间及其上的随机过程来表达复杂随机徘徊．

3*．试构造一个概率空间及其上坐标过程$\{\xi_t,t\in R^+\}$，使得$\xi_t$是独立增量的正态过程：即满足

1) 对任意的$t_1<t_2<\cdots<t_n$，$\xi_{t_2}-\xi_{t_1},\cdots,\xi_{t_n}-\xi_{t_{n-1}}$相互独立；

2) $\xi_t-\xi_s(t>s)$的分布是均值为 0，方差为$t-s$的正态分布．

# 第二章 随机徘徊与布朗运动

在第一章§1.3中我们介绍了一维对称简单随机徘徊,本章讨论较一般的随机徘徊模型及其性质与应用.

考虑一个粒子在$d$-维空间格点(记为$Z^d$)中的随机运动:粒子每隔单位时间相互独立地走一步,每步可沿任意一个坐标方向走一个单位长.设粒子沿第$k$个坐标轴正向或负向走一个单位的概率分别是$p_k$与$q_k$,则

$$\sum_{k=1}^{d}(p_k+q_k)=1.$$

于是第$k$步粒子的位移是一个随机变量:

$$x_k(\omega)=\begin{cases}\underline{e}_1, & \text{概率为 } p_1;\\ -\underline{e}_1, & \text{概率为 } q_1;\\ \cdots & \cdots\\ \underline{e}_d, & \text{概率为 } p_d;\\ -\underline{e}_d, & \text{概率为 } q_d,\end{cases} \tag{2.1}$$

其中$\underline{e}_1,\cdots,\underline{e}_d$分别表示沿第$1,\cdots,d$个坐标方向的单位向量,$p_k$,$q_k\geqslant 0(0\leqslant k\leqslant d)$.

为了要容纳所有的(无限个)相互独立的$x_k(\omega)(k=1,2,\cdots)$,我们只要取$\Omega=\{(\omega_0,\omega_1,\cdots,\omega_n,\cdots);\omega_0\in Z^d,\omega_k\in\{\pm\underline{e}_1,\pm\underline{e}_2,\cdots,\pm\underline{e}_d\},k\geqslant 1\}$,而$\mathscr{F}$为包含有限个时间任意限定条件的一切事件的最小事件体,并定义乘积概率$P$:

$$P(\omega;\omega_0\in A_0,\omega_1\in B_1,\cdots,\omega_n\in B_n)$$

$$=P_0(A_0)\prod_{k=1}^{n}P(B_k), \tag{2.2}$$

其中$n\geqslant 0$,$P_0(\cdot)$是$Z^d$上任何一个概率分布,它代表随机徘徊的

初始值 $\omega_0$ 的分布；$B_k \subset \{\pm e_1, \pm e_2, \cdots, \pm e_d\}$，$P(B_k)$是 $x_k(\omega)$的分布(2.1)：

$$\begin{Bmatrix} e_1, & -e_1, & \cdots, & e_d, & -e_d \\ p_1, & q_1, & \cdots, & p_d, & q_d \end{Bmatrix} \tag{2.3}$$

所决定的概率，于是在$(\Omega, \mathscr{F}, P)$中的坐标过程即可取为

$$x_k(\omega) = \omega_k \quad (k = 1, 2, \cdots).$$

令

$$\xi_n(\omega) = \omega_0 + \sum_{k=1}^{n} x_k(\omega), \tag{2.4}$$

就是我们想得到的简单随机徘徊的数学表达.

在概率论与应用的文献中随机徘徊还可以有更广义的理解：形如(2.4)的独立同分布随机变量的和($\omega_0 = 0$). 也就是说一般的随机徘徊可以不局限于每次沿坐标轴方向走一个单位长(如可以有各种跳动). 在本章中我们主要讨论简单随机徘徊的性质与应用. 简单随机徘徊是一个很典型的随机运动模型，在生物、物理、化学、计算方法等许多方面都有大量应用.

## §2.1 简单随机徘徊的分布与首次返回(或离开)时间

### 1. 简单随机徘徊的分布

设简单随机徘徊在时刻 $m$ 的位置为 $x$，在时刻 $n+m$ 的位置为 $y$，则经过 $n$ 个单位时间它的位移是 $y-x$，这个事件的概率是多项分布：

$$p_n(y-x) = \sum \begin{Bmatrix} n \\ k_1, h_1; k_2, h_2; \cdots; k_d, h_d \end{Bmatrix} p_1^{k_1} q_1^{h_1} p_2^{k_2} q_2^{h_2} \cdots p_d^{k_d} q_d^{h_d}, \tag{2.5}$$

其中和号取遍满足以下条件的各项：

$$\begin{cases} k_1 + h_1 + k_2 + h_2 + \cdots + k_d + h_d = n, \\ k_i - h_i = y_i - x_i \quad (i = 1,2,\cdots,d), \end{cases}$$

而 $k_i$ 与 $h_i$ 分别表示在 $n$ 步中沿 $e_i$ 的正向与负向所走的步数；$\binom{n}{m_1,m_2,\cdots,m_{2d}}$ 表示将 $n$ 个元素分成 $2d$ 堆，各有 $m_1,\cdots,m_{2d}$ 个元素的分法数，它等于 $\frac{n!}{m_1!m_2!\cdots m_{2d}!}$. 简单随机徘徊 $\xi=\{\xi_n;n=0,1,\cdots\}$ 的有限维分布为

$$P(\xi_{n_1}=x_1,\cdots,\xi_{n_l}=x_l)$$
$$=\sum_{x_0\in Z^d} p_0(x_0)p_{n_1}(x_1-x_0)p_{n_2-n_1}(x_2-x_1)\cdots p_{n_l-n_{l-1}}(x_l-x_{l-1}). \tag{2.6}$$

特别，当 $d=1,\omega_0=0$ 时

$$p_n(y-x)=\begin{cases} \dfrac{n!}{\left(\dfrac{y-x+n}{2}\right)!\left(\dfrac{n-y+x}{2}\right)!}p_1^{\frac{y-x+n}{2}}q_1^{\frac{n-y+x}{2}}, \\ \qquad \text{当 } |y-x|\leqslant n, \text{且 } y-x+n \text{ 为偶数时；} \\ 0, \quad \text{其他}, \end{cases} \tag{2.7}$$

它正是二项分布.

当 $d=2$，$\xi$ 是平面上的随机徘徊，$y-x=(y_1-x_1,y_2-x_2)$，且 $n+(y_1-x_1)+(y_2-x_2)$ 为偶数时

$$p_n(y-x)$$
$$=\sum_{k=0}^{n}\frac{n!p_1^kq_1^{k-y_1+x_1}p_2^{\frac{n+y_1-x_1+y_2-x_2}{2}-k}}{k!(k-y_1+x_1)!\left(\dfrac{n+y_1-x_1+y_2-x_2}{2}-k\right)!}$$
$$\times\frac{q_2^{\frac{n+y_1-x_1-y_2+x_2}{2}-k}}{\left(\dfrac{n+y_1-x_1-y_2+x_2}{2}-k\right)!}. \tag{2.8}$$

而在其他情形，则 $p_n(y-x)=0$. 又若随机徘徊为上下、左右均对

称的，则 $p_1=q_1=p_2=q_2=\frac{1}{4}$，那么当 $n+(y_1-x_1)+(y_2-x_2)$ 为偶数时

$$
\begin{aligned}
p_n(y-x) &= \sum_{k=0}^{n} \frac{n!\,4^{-n}}{k!(k-y_1+x_1)!\left(\frac{n+y_1-x_1+y_2-x_2}{2}-k\right)!} \\
&\quad \times \frac{1}{\left(\frac{n+y_1-x_1-y_2+x_2}{2}-k\right)!}.
\end{aligned} \tag{2.9}
$$

## 2. 1-维简单随机徘徊的反射原理与首中时

本节中，我们只讨论 1-维简单随机徘徊，即设 $d=1$，并设 $\omega_0=0$（从原点出发）．考虑 $\xi$ 首达 $y(y>0)$ 的时间：

$$T_y(\omega)=\inf\{n;\xi_n(\omega)=y\}=\inf\{n;\xi_n(\omega)\geqslant y\},$$

也称 $T_y$ 为首中时．现在来求 $T_y$ 的分布，即求 $P_0(T_y=N)$，这里 $P_0$ 的下标 0 表示从 0 点出发．显然

$$P_0(T_y=N)=0 \quad (N<y),$$

$$P_0(T_y=y)=p^y.$$

对于 $N>y$ 的情况，注意到

$$\{\omega;T_y(\omega)=N\}=\{\omega;\xi_N(\omega)=y,\xi_n(\omega)\neq y,\forall\, n<N\},$$

此事件是 $\{\omega;\xi_N(\omega)=y\}$ 的一部分．当 $N+y$ 为奇数时，它是不可能事件；而当 $N+y$ 为偶数时，它由若干个概率为 $p^{\frac{N+y}{2}}q^{\frac{N-y}{2}}$ 的等概率事件组成．设在 $\{\omega;\xi_N(\omega)=y,T_y(\omega)<N\}$ 中有 $l'$ 个上述等概事件，那么

$$
\begin{aligned}
P_0(T_y(\omega)=N) &= P_0(\xi_N(\omega)=y)-l'p^{\frac{N+y}{2}}q^{\frac{N-y}{2}} \\
&= \left(\binom{N}{(N+y)/2}-l'\right)p^{\frac{N+y}{2}}q^{\frac{N-y}{2}},
\end{aligned}
$$

这 $l'$ 个事件又分为两组，分别满足条件

a) $\xi_{N-1}(\omega)=y+1$；

b) $\xi_{N-1}(\omega)=y-1$，但在 $N-1$ 前 $\xi$ 曾到过 $y$.

a)中的事件 $\{\omega:\xi_{N-1}(\omega)=y+1\}$ 共 $\binom{N-1}{(y+N)/2}$ 个概率为 $p^{\frac{N+y}{2}}q^{\frac{N-y}{2}}$ 的等概率事件，而 b)中的事件由于下面的反射原理(引理 2.2)也都与 $\{\omega:\xi_{N-1}(\omega)=y+1\}$ 中的上述等概事件一一对应，因而也有 $\binom{N-1}{(y+N)/2}$ 个，从而

$$l'=2\binom{N-1}{(y+N)/2}.$$

于是当 $y>0$，

$$P_0(T_y=N)=\left[\binom{N}{(y+N)/2}-2\binom{N-1}{(y+N)/2}\right]p^{\frac{N+y}{2}}q^{\frac{N-y}{2}}$$

$$=\frac{y}{N}\binom{N}{(N+y)/2}p^{\frac{N+y}{2}}q^{\frac{N-y}{2}}.$$

综上结果，我们有以下定理：

**定理 2.1** 1 维简单随机徘徊，当 $\omega_0=0$ 时，首达 $y$ 的时间 $T_y$ 的分布是

$$P_0(T_y=N)=\begin{cases}\binom{N}{(y+N)/2}\frac{|y|}{N}p^{\frac{N+y}{2}}q^{\frac{N-y}{2}}, \\ \qquad |y|\leqslant N, \text{且 } y+N \text{ 为偶数时}; \\ 0, \quad |y|>N, \text{或其他情形}.\end{cases} \tag{2.10}$$

**引理 2.2(反射原理)** 由 0 出发，在时刻 $N-1$ 到 $y-1$，其间曾到过 $y$ 的全体轨道集 $A$ 与由 0 出发，时刻 $N-1$ 时到达 $y+1$ 其间曾到过 $y$ 的全体轨道集 $B$ 一一对应.

**证明** 如果将 $A$ 中的每一轨道，从其首达 $y$ 时刻 $\tau$ 起到时刻 $N-1$，对 $\xi=y$ 作一个镜面反射(见图 2.1)，就得到一个时刻 $N-1$ 时 $\xi_{N-1}=y+1$ 的轨道，这就建立了 $A$ 与 $B$ 间的一一对应.

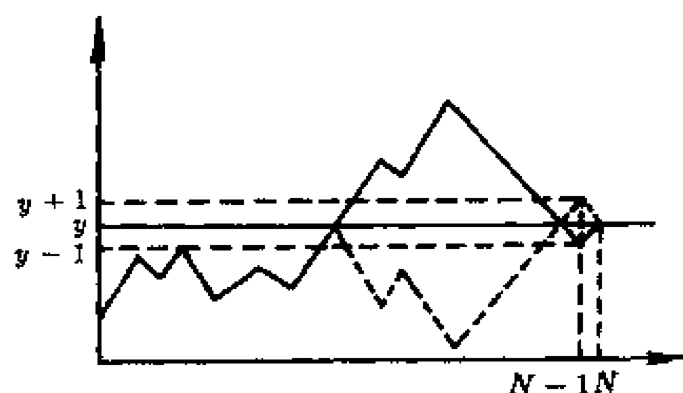


图 2.4

### 3. 赌徒破产模型

设某赌徒有赌本 $x(x\geqslant 1)$元，每赌一次独立地以 $p$ 的概率赢 1 元，以 $q=1-p$ 的概率输 1 元，若输光则必须停止. 若此人下决心在赢到本利共 $b(b>x)$元时即停止，求赌徒的破产(输光)的概率. 如果将到第 $n$ 次赌完时此赌徒在 $n$ 次全部输赢的代数和记为 $\tilde{\xi}_n$，令 $\xi_0=x$，$\xi_n=\xi_0+\tilde{\xi}_n$，则 $\xi=\{\xi_n;\ n\geqslant 0\}$ 是一个简单随机徘徊. 令

$$T_0(\omega)=\inf\{n;\xi_n(\omega)=0\},$$
$$T_b(\omega)=\inf\{n;\xi_n(\omega)=b\}.$$

它们分别是赌徒输光或主动停止赌博的时间，而他输光的概率是:

$$\varphi(x)\triangleq P_x(T_0(\omega)<T_b(\omega))\quad (P_x \text{ 表示 } \xi_0=x).$$

注意到由全概公式我们可得:

$$\varphi(x)=p\cdot\varphi(x+1)+q\cdot\varphi(x-1),\qquad (2.11)$$

并有边界条件

$$\varphi(0)=1,\qquad \varphi(b)=0.\qquad (2.11)'$$

于是，可由此解得 $\varphi(x)$. 这个问题同样可考虑得更为一般，即左端点可以不是 0，而为一个值 $a>-\infty$而且 $a\leqslant x\leqslant b$. 令

$$\varphi(x)\triangleq P_x(T_a(\omega)<T_b(\omega)).$$

它仍满足方程(2.11)与边界条件

$$\varphi(a) = 1, \qquad \varphi(b) = 0.$$

为解(2.11),注意到它等价于

$$p(\varphi(x+1) - \varphi(x)) = q(\varphi(x) - \varphi(x-1)). \qquad (2.12)$$

当 $p>0$,且 $p\neq\frac{1}{2}$时

$$\varphi(k+1) - \varphi(k) = \left(\frac{q}{p}\right)^{k-a}(\varphi(a+1) - \varphi(a)).$$

于是就可解得

$$\begin{aligned}\varphi(x) &= \sum_{k=a}^{x-1}(\varphi(k+1) - \varphi(k)) + \varphi(a) \\ &= \sum_{k=a}^{x-1}\left(\frac{q}{p}\right)^{k}(\varphi(a+1) - \varphi(a)) + \varphi(a) \\ &= \left(1 - \left(\frac{q}{p}\right)^{x-a}\right)\alpha + 1,\end{aligned}$$

其中 $\alpha$ 是与 $x$ 无关的待定常数.再由 $\varphi(b)=0$,就得到:当 $p\neq q$ 时有

$$\alpha = -\left(1 - \left(\frac{q}{p}\right)^{b-a}\right)^{-1}.$$

从而得到:当 $p\neq q, a\leqslant x\leqslant b$,有

$$\varphi(x) = \left[\left(\frac{q}{p}\right)^{x-a} - \left(\frac{q}{p}\right)^{b-a}\right]\Big/\left[1 - \left(\frac{q}{p}\right)^{b-a}\right]. \qquad (2.13)$$

当 $p=q=\frac{1}{2}$,由(2.12),我们有

$$\varphi(x) = (x-a)(\varphi(a+1) - \varphi(a)) + \varphi(a) \triangleq (x-a)\alpha + 1.$$

再由 $\varphi(b)=0$,就解得

$$\alpha = -(b-a)^{-1}, \quad \varphi(x) = (b-x)/(b-a). \qquad (2.13)'$$

从上面的结果,可以看出,在公平赌博$\left(p=\frac{1}{2}\right)$中,如果开始的赌本为 $x$ 元,希望赢到本利共 $b$ 元后才停下,那么,他输光的概率是 $(b-x)/b$.可见,如果赌本相对于希望赢的钱很少,即 $x/b\sim 0$,那么他以几乎为 1 的概率是要以输光告终的.当赌徒希望赢到本利

和为 $k$ 倍于赌本的钱方罢休，那么他能达到他的希望的概率只是 $1/k$——越贪心就越可能输光.

### 4. 带吸收壁的随机徘徊

设 $\xi \triangleq \{\xi_n; n\geqslant 0\}$ 是 1-维简单随机徘徊，令

$$\tau = T_a(\omega) \wedge T_b(\omega) = \inf\{n; \xi_n(\omega) \notin [a,b)\},$$

$$\eta_n(\omega) = \xi_{n\wedge\tau(\omega)},$$

意思是设想在 $a,b$ 两点分别设一个吸收壁，粒子到达 $a$ 或 $b$ 后就被吸收而永远停留在那里. 因此，$\eta \triangleq \{\eta_n; n\geqslant 0\}$ 就称为两端带吸收壁的随机徘徊. 若只在一个点 $a$ 设置吸收壁，即令

$$\bar{\eta}_n(\omega) \triangleq \xi_{n\wedge T_a(\omega)}(\omega) \quad (\eta \geqslant 0),$$

那么 $\bar{\eta} \triangleq \{\bar{\eta}_n; n\geqslant 0\}$ 就是一端吸收壁的随机徘徊. 其实前面的赌徒破产问题中希望赢利有限与无限两种情况就分别对应于这两种吸收壁的随机徘徊.

吸收时间 $T_a(\omega)$ 的分布其实已由(2.10)式给出：

$$P_x(T_a(\omega) = N) = P_0(T_{a-x}(\omega) = N),$$

这是因为注意到简单随机徘徊是对空间平移不变的，因而将原空间向左平移 $x$ 就得到了上述结果. 原则上

$$P_x(T_a(\omega) < +\infty) = \sum_{N=|a-x|}^{\infty} P_0(T_{a-x}(\omega) = N),$$

但上面右端的无穷级数求和需要较高的分析技巧. 这里我们绕道由(2.13)与(2.13)′来求它. 注意到上式等于

$$\lim_{b\to+\infty} P_x(T_a(\omega) < T_b(\omega))$$

(因为 $P_x(\lim\limits_{b\to+\infty} T_b(\omega) = +\infty) \geqslant P_x(T_b(\omega) \geqslant b-x) = 1$),

$$P_x(T_a(\omega) < +\infty) = P_x(T_a(\omega) < \lim_{b\to+\infty} T_b(\omega))$$

$$= \begin{cases} \lim\limits_{b\to+\infty}\left[\left(\dfrac{q}{p}\right)^{x-a} - \left(\dfrac{q}{p}\right)^{b-a}\right]\left[1-\left(\dfrac{q}{p}\right)^{b-a}\right]^{-1} = \left(\dfrac{q}{p}\right)^{x-a}, & p>q; \\ 1, & p\leqslant q, \end{cases}$$

这里 $p>q$ 当且仅当 $p>\frac{1}{2}$.

## §2.2 Brown 运动

1827 年英国生物学家 Brown 在显微镜下观测液面上的花粉，发现花粉的微粒作着高度不规则的运动. 但是直到 19 世纪末，人们才清楚这种奇怪的现象是由于花粉微粒受到大量液体分子的无规则碰撞造成的. Einstein 对此现象首次作出理论的、量化的分析. 本节将介绍 Einstein 分析的梗概. 另一方面，Brown 运动又可以作为随机徘徊的"极限"纯数学地得到. 这个认识 Brown 运动的途径不仅简单而直观，也为 Brown 运动及与它有关的随机过程及随机微分方程近似计算提供了基础. 这个作法的直观思想由法国数学家 Bachellier 首先提出，此后 Kolmogorov 关于马氏过程类型的随机过程理论很大程度上是基于这种想法的. 本节也给出这种作法的基本思想.

Paul Levy 从 1910 起数十年的工作，对 Brown 运动的研究有着深远的影响. 他的著作《Processes stochastiques et mouvement brownien》(1948 年第一版，1965 年第二版)至今仍对这方面的研究工作有许多启示与参考价值. 在 Levy 早期工作的基础上 Wiener 第一个(1923 年)严格地给出 Brown 运动的数学定义：构造出了一个概率空间(Wiener 空间)及其上的随机过程来刻画 Einstein 的物理严格意义下的 Brown 运动. 因而 Brown 运动也叫做 Wiener 过程. Wiener 的论文"Differential space", J. Math and phys. 2, 131－174 是 Brown 运动研究的里程碑. 可以这样说：由 Einstein 首创的 Brown 运动的数学模型，由 Levy 与 Wiener 大大地发展深化了. 至今，由于大量的数学家与自然科学家的工作，Brown 运动及其泛函的研究不断深入发展，它已成为随机过程的两大基石之一. 它不仅渗透到偏微分方程、调和分析、计算方法、控

制等各数学领域，而且在生物、化学、物理、力学、工程、经济管理、金融等学科中 Brown 运动也成为不可缺少的研究工具. 它是"噪声"与"涨落"等随机现象的典型，并提供处理的参考模式.

## 1. Brown 运动分布的推导

本节得到 Brown 运动分布的分析方法本质上是 Einstein 的方法，但为了避免解偏微分方程，我们对分布密度计算改用其特征函数.

按照 Einstein 的分析，一个 Brown 运动是表示一个随机运动着的粒子在时间$[0,t]$上的位移 $B_t$，显然 $B_0=0$. 粒子运动应有以下性质：

1）粒子位移的各分量都相互独立. 我们不妨只考虑其一个分量，仍记为$\{B_t\}$. 而且对任意互不相交的区间$[s_1,t_1)$，$[s_2,t_2)$，…，$[s_n,t_n)$有 $B_{t_1}-B_{s_1}$，$B_{t_2}-B_{s_2}$，…，$B_{t_n}-B_{s_n}$都相互独立.

2）运动的统计规律对空间是对称的，因而

$$EB_t = 0.$$

3）$B_{t+h}-B_h$ 的分布与 $h$ 无关，并且 $\sigma(t)\triangleq E(B_{s+t}-B_s)^2$ 存在，而且是 $t$ 的连续函数. 设 $B_t$ 的分布密度（我们假定密度存在，否则就要用测度论工具）是 $\rho(t,x)$，令

$$\varphi(x,\lambda) = \int e^{i\lambda x}\rho(t,x)dx$$

与

$$\sigma(t) = E(B_t^2) = \int x^2\rho(t,x)dx \quad (EB_t = 0),$$

于是由上述性质就可以知道

$$\begin{aligned}\sigma(t+s) &= E(B_{t+s}^2) = E(B_{t+s}-B_t+B_t-B_0)^2 \\ &= E(B_{t+s}-B_t)^2 + E(B_t^2) \qquad (\text{用 } B_0 = 0 \text{ 与(1)}) \\ &= \sigma(s)+\sigma(t) \qquad (\text{因 } B_{t+s}-B_t \text{ 分布与 } t \text{ 无关}).\end{aligned}$$

又由于 $\sigma(t)$对 $t$ 连续，由数学分析的分析不难得到

$$\sigma(t) = D \cdot t,$$

其中 $D$ 是一个待定常数，它是单位时间中粒子平方位移的均值，也称扩散常数. 在分子运动学物理中它是：

$$D = \frac{RT}{Nf}. \tag{2.15}$$

其中 $R$ 是由分子的特性决定的一个普适常数，$T$ 是绝对温度，$N$ 是阿弗加德罗（Avogadro）常数，$f$ 是摩擦系数.（2. 15）可由理论分析得到，它与实验很好地符合. 这里我们略去这一物理推导.

再利用 1）与 3），我们得到：

$$\begin{aligned}
\varphi(t+s,\lambda) - \varphi(t,\lambda) &= E(e^{i\lambda B_{t+s}}) - E(e^{i\lambda B_t}) \\
&= E(e^{i\lambda B_t}(e^{i\lambda(B_{t+s}-B_t)} - 1)) \\
&= E((e^{i\lambda(B_t-B_0)})(e^{i\lambda(B_{t+s}-B_t)} - 1)) \\
&= E(e^{i\lambda(B_t-B_0)})E(e^{i\lambda(B_{t+s}-B_t)} - 1) \quad (\text{独立性}) \\
&= E(e^{i\lambda B_t})E(e^{i\lambda(B_s-B_0)} - 1) \\
&= \varphi(t,\lambda)E(e^{i\lambda B_s} - 1).
\end{aligned} \tag{2.16}$$

注意到估计式：

$$e^{iv} = 1 + iv - \frac{v^2}{2} + o(v^2),$$

我们有

$$E(e^{i\lambda B_s} - 1) = i\lambda EB_s - \frac{1}{2}\lambda^2 EB_s^2 + o(EB_s^2)$$

$$= -\frac{1}{2}\lambda^2\sigma(s) + o(s).$$

对(2. 16)两端除以 $s$ 并令 $s \to 0$ 便得到：

$$\frac{\partial\varphi}{\partial t}(t,\lambda) = -\frac{1}{2}\lambda^2\varphi(t,\lambda)D,$$

这是对 $t$ 的一个常微分方程，解这个方程得到

$$\varphi(t,\lambda) = \varphi(0,\lambda)e^{-\frac{1}{2}\lambda^2 tD}.$$

再注意到

$$\varphi(0,\lambda) = E(\mathrm{e}^{\mathrm{i}\lambda B_0}) = 1,$$

就有

$$\varphi(t,\lambda) = \mathrm{e}^{-\frac{1}{2}\lambda^2 Dt} = \int \mathrm{e}^{\mathrm{i}\lambda x} \frac{\mathrm{e}^{-\frac{x^2}{2Dt}}}{\sqrt{2\pi Dt}} \mathrm{d}x. \qquad (2.17)$$

可见 $B_t$ 的分布是 $N(0,Dt)$. 为了方便起见,数学上我们取适当单位使 $D=1$.

于是由 1)就得到:$B_{t_1}, B_{t_2}-B_{t_1}, \cdots, B_{t_n}-B_{t_{n-1}}$ 的联合分布是各正态分布的乘积,因而就得到 Brown 运动在任意有限个时刻的联合分布是:

$$P(\omega: B_{t_1} < x_1, \cdots, B_{t_n} < x_n)$$

$$= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \frac{\mathrm{e}^{-\left\{\frac{u_1^2}{2t_1} + \frac{(u_2-u_1)^2}{2(t_2-t_1)} + \cdots + \frac{(u_n-u_{n-1})^2}{2(t_n-t_{n-1})}\right\}}}{(2\pi)^{n/2}\sqrt{t_1(t_2-t_1)\cdots(t_n-t_{n-1})}} \mathrm{d}u_1 \cdots \mathrm{d}u_n, \qquad (2.18)$$

于是,一般地 Brown 运动可定义为:

**定义 2.1** Brown 运动是一个随机过程 $\boldsymbol{B}=\{B_t: t\geqslant 0\}$,满足:

*B*.1) 独立增量,即对任意互不相交的 $n$ 个区间 $[s_1,t_1), [s_2,t_2), \cdots, [s_n,t_n)$,其上 $\boldsymbol{B}$ 的增量 $B_{t_1}-B_{s_1}, B_{t_2}-B_{s_2}, \cdots, B_{t_n}-B_{s_n}$ 都相互独立;

*B*.2) $B_{t+h}-B_h$ 遵从均值为 0,方差为 $t$ 的正态分布 $N(0,t)$;

*B*.3) 对每一个 $\omega$, $B_t(\omega)$ 是 $t$ 的连续函数.

事实上,*B*.1)与 *B*.2)已完全决定了 $B$ 的任意有限维分布,也就决定了 $B$ 的分布. 要求 *B*.3)成立是由于 $B_t$ 表示粒子在 $[0,t]$ 的位移,因而应对 $t$ 连续. 这一性质不能由 $B$ 的分布决定,但是可以证明满足 *B*.1),*B*.2)的随机过程,在其概率空间 $\Omega$ 中除去一个零概率集外的任一轨道 $\omega$, $B_t(\omega)$ 都是 $t$ 的连续函数(见本章附录).

## 2. Brown 运动作为随机徘徊的极限

另一个简单而有用的对 Brown 运动定义的理解是:把它看成

某种意义下对称随机徘徊的某种极限过程. 这种理解的好处是：它不仅简单而直观，并且可为连续参数的随机过程 Brown 运动的数值计算的离散化提供参考. 考虑一个对称的无限制的随机徘徊 $\xi=\{\xi_n;n\geqslant 0\}$，如果它每走一步用的单位时间是 $\Delta t$，步长是 $\Delta x$，那么它走一步，平均位移是 0，其方差是 $(\Delta x)^2$. 若在时间 $t$ 中 $\xi$ 近似走了 $n$ 步，即 $t\approx n\Delta t$，它的平均位移仍为 0，方差是 $n(\Delta x)^2=\frac{t}{\Delta t}(\Delta x)^2$. 于是当 $\Delta x\to 0,\Delta t\to 0$ 时，一般来说下式（时间 $t$ 中的位移）不会得到有意义的极限：

$$\Delta x\cdot\xi_n(\omega)=\sum_{k=1}^{n}x_k(\omega)\Delta x.$$

当 $\Delta x$ 取 $\sqrt{D\Delta t}$ 时，由 DeMoivre（德莫弗）或中心极限定理就得到：

$$\lim_{n\to\infty}P\left\{\frac{1}{\sqrt{n}}\sum_{k=1}^{n}x_k(\omega)<x\right\}=\int_{-\infty}^{x}\frac{e^{-\frac{u^2}{2t}}}{\sqrt{2\pi t}}\mathrm{d}u$$

（这里我们不失一般性地取 $D=1$）. 于是，作为随机徘徊的极限，Brown 运动在任意一个长为 $t$ 的时间区间中的位移应为正态分布 $N(0,t)$ 的随机变量. 此即在上段中 Brown 运动定义中的 $B.2$)，又由于随机徘徊就是独立增量的，它的极限当然也是独立增量的. 这样我们就又可以把 Brown 运动看成是当 $\frac{(\Delta x)^2}{\Delta t}=D(=1)$ 时，$\Delta t\to 0$ 意义下的极限（这个极限过程的确切含义就是 § 2.3 中的不变原理）.

### 3. Brown 运动的简单性质

**定理 2.2** 设随机过程 $\{\xi_t;t\geqslant 0\}$ 满足 $\xi_0=0$，则它具有 Brown 分布当且仅当它是 Gauss 系（即任意有限维（联合）分布都是 Gauss 分布），而且 $E\xi_t=0$，$E(\xi_t\xi_s)=t\wedge s$.

**证明** 必要性：显然，具有 Brown 分布的随机过程是 Gauss 系，且 $E\xi_t=0$. 又

$$E(\xi_t\xi_s)=\frac{1}{2}E(\xi_t^2+\xi_s^2-(\xi_t-\xi_s)^2)$$
$$=\frac{1}{2}(t+s-|t-s|)=t\wedge s.$$

充分性： 当条件成立时，我们有

$$E(\xi_t-\xi_s)=0,$$
$$E(\xi_t-\xi_s)^2=t+s-2(t\wedge s)=|t-s|,$$
$$E(\xi_{t_1}-\xi_{s_1})(\xi_{t_2}-\xi_{s_2})=t_1-s_1-(t_1-s_1)$$
$$=0\quad(\forall\, s_1<t_1\leqslant s_2<t_2).$$

可见 $\xi_t-\xi_s\sim N(0,t-s)$. 由于 Gauss 系中有限个随机变量的不相关性就是独立性，因而对于 $\forall\, s_1<t_1\leqslant s_2<t_2\leqslant s_3<\cdots\leqslant s_n<t_n$，我们有 $\xi_{t_1}-\xi_{s_1},\cdots,\xi_{t_n}-\xi_{s_n}$ 相互独立，这就说明了 $\xi$ 具有 Brown 分布.

**定理 2.3** 设 $\{B_t:t>0\}$ 是 Brown 运动，那么以下各过程均为 Brown 运动：

(1) $\{-B_t:t\geqslant 0\}$；

(2) $\{B_{t+t_0}-B_{t_0}:t\geqslant 0\}$；

(3) $\left\{\frac{1}{\sqrt{\lambda}}B_{\lambda t}:t\geqslant 0\right\}\quad(\lambda>0)$；

(4) $\{tB_{\frac{1}{t}}I_{(t>0)}:t\geqslant 0\}$（其中 $I_A$ 表示 $A$ 的示性函数，$t=0$ 时 $tB_{\frac{1}{t}}I_{t>0}$ 理解为 0）；

(5) $\{B_{T-t}-B_T:0\leqslant t\leqslant T\}$.

**证明** 由定理 2.2 及以下各式，(1)～(5)的结论显然成立：

$$E(-B_t)=E(B_{t+t_0}-B_{t_0})=E\left(\frac{1}{\sqrt{\lambda}}B_{\lambda t}\right)=E(tB_{\frac{1}{t}})$$
$$=E(B_{T-t}-B_T)=0,$$
$$E((-B_t)(-B_s))=E(B_tB_s)=t\wedge s,$$
$$E(B_{t+t_0}-B_{t_0})(B_{s+t_0}-B_{t_0})$$
$$=(t+t_0)\wedge(s+t_0)-t_0-t_0+t_0=t\wedge s.$$

$$E\left(\frac{1}{\sqrt{\lambda}}B_{\lambda t}\right)\left(\frac{1}{\sqrt{\lambda}}B_{\lambda s}\right)=\frac{1}{\lambda}((\lambda t)\wedge(\lambda s))=t\wedge s,$$

$$E(tB_{\frac{1}{t}}sB_{\frac{1}{s}})=ts\left(\frac{1}{t}\wedge\frac{1}{s}\right)=t\wedge s,$$

$$\begin{aligned}E((B_T-B_t)(B_T-B_s))&\\=(T-t)\wedge(T-s)&=(T-t)-(T-s)+T\\=T-s\vee t+t+s&-T-t\wedge s.\end{aligned}$$

## § 2.3　不变原理与 Brown 运动的性质

### 1. 不变原理

在上节中,我们看到当步长 $\Delta x=\sqrt{\Delta t}$ 时,随机徘徊 $\xi_{[\frac{t}{\Delta t}]}\Delta x=\xi_{[nt]}/\sqrt{n}\left(取\frac{1}{\Delta t}=n\right)$ 以 Brown 运动为极限. 其实这个事实可以推广到一般的随机徘徊. 即有以下的**泛函中心极限定理**,意思是:相应于随机变量的中心极限定理,这里是考虑随机过程的相应定理. 另一方面,由于极限的分布与生成随机徘徊的随机变量无关,因而此定理又称**不变原理**.

**定理 2.5(泛函中心极限定理)**　设 $\{x_n,n\geqslant 1\}$ 是 i.i.d.(独立同分布)序列,$Ex_n=0$,方差为 1,则当 $n\to+\infty$ 时,$\{x_t^{(n)}:t\geqslant 0\}$ 或 $\{\tilde{x}_t^{(n)}:t\geqslant 0\}$ 也按分布收敛到标准 Brown 运动 $\boldsymbol{B}$,其中

$$x_t^{(n)}=\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}x_k(\omega)=\xi_{[nt]}/\sqrt{n}\quad\left(其实\ x_{k/n}^{(n)}=\frac{1}{\sqrt{n}}\xi_k\right),$$

而

$$\tilde{x}_t^{(n)}=\frac{\xi_{[nt]}}{\sqrt{n}}+(nt-[nt])\frac{x_{[nt]+1}}{\sqrt{n}}\quad(t\geqslant 0)$$

(其实,$\tilde{x}_t^{(n)}$ 是将点 $\{x_{\frac{m}{n}}^{(n)}:m=0,1,\cdots\}$ 用折线联结而得到的函数).

定理中"按分布收敛"的意义需要有测度论的知识才能严格讲

清楚.这里我们粗略地解释 $\tilde{x}^{(n)}$ 按分布收敛到 $\boldsymbol{B}$:对任意连续函数空间 $C[0,+\infty)$ 上的有界连续泛函 $f$,随机变量 $f(\tilde{x}^{(n)})$ 的数学期望收敛到 $f(\boldsymbol{B})$ 的数学期望(严格的定义见附录).至于 $x^{(n)}$ 的收敛性虽不能按上面解释,但仍可理解为任意样本的有界连续泛函 $f$,有 $f(x^{(n)})$ 的数学期望收敛到 $f(\boldsymbol{B})$ 的数学期望.一般地,它蕴含了 $x^{(n)}$ 的有限维分布收敛到 $\boldsymbol{B}$ 的有限维分布.

对于随机向量 $x_k(\omega)$,如果 $x_k(\omega)=(x_k^{(1)},x_k^{(2)},\cdots,x_k^{(d)})$ 而且 $Ex_k^{(i)}x_k^{(j)}=0(i\neq j)$,那么定理 2.3 对于随机向量依然适用,只是这时极限是一个 $d$-维的 Brown 运动了.所谓 $d$-维 Brown 运动,指一个 $d$-维向量值的随机过程 $\boldsymbol{B}=\{B_t;t\geqslant 0\}$,$B_t=(B_t^{(1)},\cdots,B_t^{d})$,$\{B_t^{(1)}\},\cdots,\{B_t^{d}\}$ 是相互独立的一维 Brown 运动.这也等价于下面的定义:

**定义 2.1′** 将定义 2.1 中 $B.2)$ 改为

$B.2)'$ $B_{t+s}-B_s$ 是遵从 0 均值,$t\boldsymbol{I}$ 为协方差阵的正态分布(其中 $\boldsymbol{I}$ 表示 $d$-阶单位阵)的随机向量.

## 2. 非对称的随机徘徊的极限——带漂移的 Brown 运动

对非对称的简单随机徘徊,或 $EX_n=\mu\neq 0$ 的一般随机徘徊,$\xi_{[nt]}/\sqrt{n}$ 的极限是什么样的呢?事实上考虑

$$\tilde{\xi}_{[nt]}\triangleq\xi_{[nt]}-[nt]\mu,$$

则由定理 2.5 有 $\tilde{\xi}_{[nt]}/\sqrt{n}$ 收敛到 Brown 运动 $B_t$,那么当 $\mu=(p-q)\sim\frac{b}{\sqrt{n}}$ 时有极限:

$$\lim_{n\to\infty}[nt]\mu/\sqrt{n}=bt.$$

于是 $\xi_{[nt]}/\sqrt{n}=\tilde{\xi}_{[nt]}/\sqrt{n}+[nt]\mu/\sqrt{n}$ 的极限是 $B_t+bt$,记为 $\tilde{B}_t$.$\tilde{\boldsymbol{B}}\triangleq\{\tilde{B}_t;t\geqslant 0\}$ 称为一个带漂移的 Brown 运动,它具有一个常

数 $b$ 的漂移系数.

### 3. 应用于首中时及击中分布计算

利用定理 2.5 及简单随机徘徊有关首中时等结果可以很快得到 Brown 运动的相应结果. 令 $\boldsymbol{B}=\{B_t;t\geqslant 0\}$ 是标准 Brown 运动，满足 $B_0=0, y>0$，及

$$\tau_y(\omega) \triangleq \inf\{t;B_t = y\} = \inf\{t;B_t \geqslant y\}, \qquad (2.19)$$

$$\begin{aligned} T_y^{(n)}(\omega) &\triangleq \inf\{m;\bar{x}_{m/n}^{(n)} \geqslant y\} \\ &= \inf\{m;\xi_m = [y\sqrt{n}] + C\} \quad (C=0 \text{ 或 } 1). \end{aligned} \qquad (2.20)$$

**定理 2.6** $P(\tau_a<\tau_b)=\dfrac{b}{b-a}\quad (a\leqslant 0\leqslant b)$,

$$P(\tau_b < \tau_a) = \frac{-a}{b-a} \quad (a\leqslant 0\leqslant b),$$

$$P(\tau_b < +\infty) = 1 \quad (\text{任意 } b),$$

$$\begin{aligned} P(\tau_y > t) &= \sqrt{\frac{2}{\pi}}\int_0^{|y|/\sqrt{t}} \mathrm{e}^{-\frac{v^2}{2}}\mathrm{d}v \\ &= \sqrt{\frac{2}{\pi}}|y|\frac{1}{2}\int_t^{\infty}\exp(-z^2/2u)u^{-\frac{3}{2}}\mathrm{d}u. \end{aligned}$$

**直观证明** 用测度论与泛函分析的知识可以证明本定理所要求的各概率均为 $B$. 的有界连续泛函的数学期望，所以可由 $\{T_y^{(n)}\}$ 渐近地替代 $\tau_y$. 于是由(2.13)′,

1° $P(\tau_a<\tau_b)=\lim\limits_{n\to\infty}P(T_a^{(n)}<T_b^{(n)})$

$$=\lim_{n\to\infty}\frac{[b\sqrt{n}]+1}{[b\sqrt{n}]-[a\sqrt{n}]}=\frac{b}{b-a}.$$

2° $P(\tau_b<+\infty)=\lim\limits_{a\to-\infty}P(\tau_b<\tau_a)=1.$

3° $P(\tau_b<\tau_a)=1-P(\tau_a<\tau_b)=\dfrac{-a}{b-a}\quad (\tau_a\neq\tau_b).$

4° $P(\tau_y>t)=P(B_s\leqslant y;\forall\ s\leqslant t)=\lim\limits_{n\to\infty}P(x_s^{(n)}\leqslant y;\forall\ s\leqslant t)$

$$= \lim_{n\to\infty} P(\xi_{[ns]} \leqslant \sqrt{n}\, y;\forall\ s \leqslant t) = \lim_{n\to\infty} P(T_z^{(n)} > [nt])$$

$$= \lim_{n\to\infty} \sum_{m=[nt]+1}^{\infty} \frac{z}{m}\binom{m}{(m+z)/2} 2^{-m},$$

其中 $z=[\sqrt{n}\, y]+C$，$C=0$ 或 $1$，$m+z$ 为偶数，由 Sterling 公式

$$m! = \sqrt{2\pi}\, e^{-m} m^{m+\frac{1}{2}}(1+o(1)) \quad (\text{当 } m\to+\infty,\ o(1)\to 0),$$

我们有

$$\frac{z}{m}\binom{m}{(m+z)/2} 2^{-m}$$

$$= 2^{-m}\frac{z}{m}\frac{1}{\sqrt{2\pi}} e^{-m} m^{m+\frac{1}{2}} e^{\frac{m+z}{2}} \left(\frac{m+z}{2}\right)^{-\frac{m+z}{2}-\frac{1}{2}} e^{\frac{m-z}{2}}$$

$$\times \left(\frac{m-z}{2}\right)^{-\left(\frac{m-z}{2}\right)-\frac{1}{2}}(1+o(1))$$

$$= \sqrt{\frac{2}{\pi}}\, m^{-\frac{3}{2}} z \exp\left(-\frac{m+z}{2}\log\left(1+\frac{z}{m}\right)\right.$$

$$\left. -\frac{m-z}{2}\log\left(1-\frac{z}{m}\right)\right)(1+o(1))$$

$$= \sqrt{\frac{2}{\pi}}\, m^{-\frac{3}{2}} z \exp\left(-\frac{z^2}{2m}+o\left(\frac{1}{m}\right)\right)(1+o(1))$$

$$= \sqrt{\frac{2}{\pi}}\, m^{-\frac{3}{2}} z \exp\left(-\frac{z^2}{2m}\right)(1+o(1))$$

$$= \sqrt{\frac{2}{\pi}}(nt+2k)^{-\frac{3}{2}}\sqrt{n}\, y$$

$$\times \exp\left(-\frac{y^2 n}{2(2k+nt)}\right)(1+o(1)) \quad (m=[nt]+2k)$$

$$= \frac{1}{n}\sqrt{\frac{2}{\pi}}\, y\left(t+\frac{2k}{n}\right)^{-\frac{3}{2}} \exp\left(-\frac{y^2}{2\left(t+\frac{2k}{n}\right)}\right)(1+o(1)).$$

于是

$$
\begin{aligned}
P(\tau_y > t) &= \sqrt{\frac{2}{\pi}}\, y \cdot \frac{1}{2} \lim_{n\to\infty} \sum_{k=1}^{\infty} \left( t + \frac{2k}{n} \right)^{-\frac{3}{2}} \exp\left[ -\frac{y^2}{2\left( t + \frac{2k}{n} \right)} \right] \frac{2}{n} \\
&= \sqrt{\frac{2}{\pi}}\, \frac{y}{2} \int_t^{\infty} e^{-\frac{y^2}{2u}} u^{-\frac{3}{2}} \mathrm{d}u \qquad (2.21) \\
&= \sqrt{\frac{2}{\pi}} \int_0^{\frac{y}{\sqrt{t}}} e^{-\frac{v^2}{2}} \mathrm{d}v \quad \left( 令\ v = \frac{y}{\sqrt{u}} \right).
\end{aligned}
$$

对 $y<0$ 的情形,很容易得到相应结果.

## 4. 最大值分布及其均值

由于对 $y>0$ 与标准 Brown 运动 $\boldsymbol{B}$,有

$$
\{\omega: \max_{0\leqslant s\leqslant t} B_s \leqslant y\} = \{\omega: \tau_y(\omega) > t\},
$$

我们得到 $M_t = \max\limits_{0\leqslant s\leqslant t} B_s$ 的分布函数:

$$
F_M(y) \triangleq P(\max_{0\leqslant s\leqslant t} B_s \leqslant y) = \sqrt{\frac{2}{\pi}} \int_0^{\frac{y}{\sqrt{t}}} e^{-\frac{v^2}{2}} \mathrm{d}v.
$$

$M_t$ 的分布密度可由 $F_M'(y)$ 得到:

$$
f_M(y) \triangleq \begin{cases} \sqrt{\dfrac{2}{\pi t}} e^{-\frac{y^2}{2t}}, & y > 0; \\ 0, & y < 0, \end{cases}
$$

$$
EM_t = \int_0^{\infty} y \sqrt{\frac{2}{\pi t}} e^{-\frac{y^2}{2t}} \mathrm{d}y = \int_0^{\infty} e^{-u} \sqrt{\frac{2t}{\pi}} \mathrm{d}u = \sqrt{\frac{2t}{\pi}}. \qquad (2.22)
$$

## 5. 反正弦律的直观推导

令

$$
\gamma \triangleq \sup\{t: 0 \leqslant t \leqslant 1, B_t = 0\},
$$

那么 $\gamma$ 是 $\boldsymbol{B}$ 在时间 $t=1$ 前最后一次到 0 的时刻. 如果令 $B_t = B_1 -$

$B_{1-t}$，那么 $\tilde{B}=\{\tilde{B}_t;1\geqslant t\geqslant 0\}$ 就是 $B$ 在[0,1]上的时间倒逆过程，由定理 2.4 的 5)，$\tilde{B}$ 仍是标准 Brown 运动. 在 $B_1=y$ 的条件下，

$$1-\gamma(\omega)=\inf\{t:\tilde{B}_t=y\}\triangleq\tau_y;$$

由(2)式

$$P(1-\gamma(\omega)>t\,|\,B_1=y)=P(\tau_y>t\,|\,\tilde{B}_0=0)$$

$$=\int_t^{\infty}\sqrt{\frac{2}{\pi}}\,\frac{|y|}{2}e^{-\frac{y^2}{2u}}u^{-\frac{3}{2}}\mathrm{d}u,$$

即 $\gamma(\omega)$ 的条件密度是 $\frac{|y|}{\sqrt{2\pi}}e^{-\frac{y^2}{2(1-t)}}(1-t)^{-\frac{3}{2}}$，于是由全概公式

$$P_0(\gamma\in[t,t+\mathrm{d}t))=P(\gamma\in[t,t+\mathrm{d}t)\,|\,B_0=0)$$

$$=\int_{R^1}P(\gamma\in[t,t+\mathrm{d}t)\,|\,B_0=0,B_1=y)p(1,0,y)\mathrm{d}y$$

(其中 $p(t,x,y)\mathrm{d}y=P(B_t\in(y,y+\mathrm{d}y)\,|\,B_0=x)$)

$$=\int_{R^1}\frac{P_y(\tilde{\tau}_0\in(1-t-\mathrm{d}t,1-t])p(t,0,0)}{p(1,y,0)}p(1,0,y)\mathrm{d}y$$

$$=\int_{-\infty}^{+\infty}\frac{|y|}{\sqrt{2\pi}}e^{-\frac{y^2}{2(1-t)}}(1-t)^{-\frac{3}{2}}\frac{e^{-\frac{0}{2t}}}{\sqrt{2\pi t}}\mathrm{d}y\cdot\mathrm{d}t$$

$$=\frac{1}{\pi\sqrt{t(1-t)}}\mathrm{d}t,$$

即 $\gamma$ 的分布密度是 $\frac{1}{\pi\sqrt{t(1-t)}}$. $\gamma$ 的分布函数是以下反正弦律，因为

$$P(\gamma\leqslant x)=\int_0^x\frac{1}{\pi\sqrt{t(1-t)}}\mathrm{d}t=\int_0^{\sqrt{x}}\frac{2\mathrm{d}y}{\pi\sqrt{1-y^2}}\quad(y=\sqrt{t})$$

$$=\frac{2}{\pi}\arcsin\sqrt{x}.$$

## §2.4 应用——自由连接高分子链的构象分析

在高分子科学中，链构象的研究一直占中心地位. 自由连接链

是物理化学家们用简单模型来代表真实大分子的一个重要的处理方法.由此模型去发现橡皮高弹性与构象之间的关系,去解释高分子溶液的黏度等一系列问题,得到了高分子理论中许多重要概念与理论.这里我们就其基本模型与概率思想作一简要的介绍,以示概率论,特别是与本章有关的结果可如何应用于实际问题.

所谓自由连接高分子链(简称自由链)是指大分子是由一些不占体积的刚性的长为 $l$ 的键相继自由地连接起来的.因而每一个键在空间的方向与前面的各键的方向都无关.将第 $k$ 个键的方向记为 $\underline{x}_k$,于是 $\{\underline{x}_k;k\geqslant 1\}$ 是独立同分布的随机向量序列,而且沿各方向分布均匀.若大分子共 $n$ 节,那么第 $n$ 个键的末端所在位置(高分子统计中称之为末端距)是:

$$\underline{\xi}_n(\omega)=(\xi_n^{(1)}(\omega),\cdots,\xi_n^{(d)}(\omega))=\sum_{k=1}^{n}l\underline{x}_k(\omega),$$

$$\underline{x}_k(\omega)=(x_n^{(1)}(\omega),\cdots,x_n^{(d)}(\omega)),$$

其中 $d=2$ 与 3,分别表示链是在平面与空间中伸展.由对称性

$$E\underline{x}_k=0,\ E|\underline{x}_k|^2=1,\ E\xi_n^{(i)}\xi_n^{(j)}=E\Big(\sum_{k=1}^{n}x_k^{(i)}x_k^{(j)}\Big)\cdot l^2=0.$$

而令

$$C_n\triangleq E|\underline{\xi}_k|^2/nl^2\to C_\infty\quad(\text{当 } n\to+\infty).$$

在高分子构象分析中 $C_\infty$ 是表征高分子刚性的标志.例如完全伸直的链(记为 $C_\infty$),意思是各链在连接时必须完全僵直,不能弯曲;这时 $C_\infty=\lim\limits_{n\to\infty}E|\eta_n|^2/nl^2=\lim\limits_{n\to\infty}(n\cdot l)^2/nl^2=+\infty$.但自由链的均方末端距是 $E|\underline{\xi}_n|^2=nl^2E|x_k|^2=nl^2$,即 $C_\infty=1$.可见 $C_\infty$ 越大,刚性越大.$\sqrt{E|\underline{\xi}_n|^2}=(nl^2)^{1/2}$ 称为均方根末端距.本节研究自由链在空间弯曲延伸成为一个"线团"的形态的特征参数及其渐近计算($n\to+\infty$).

**1. 末端分布**

注意到 $(x_k^{(1)})^2+\cdots+(x_k^{(d)})^2=1$,那么由各方向分布均匀性得

$$1 = E((x_k^{(1)})^2 + \cdots + (x_k^{(d)})^2) = dE(x_k^{(i)})^2,$$

即 $E(x_k^{(i)})^2 = 1/d$，而且 $E(x_k^{(i)}x_k^{(j)}) = 0(i \neq j$，因为 $x_k$ 在单位球或圆上均匀分布). 当 $nl^2 \to a^2$ 时，由定理 2.5(对向量值随机过程用定理)就得到$(\xi_n^{(1)}, \cdots, \xi_n^{(d)})$的分布收敛到$(cB_1^{(1)}, \cdots, cB_1^{(d)})$的分布，其中

$$c^2 = a^2 \cdot E(x_k^{(i)})^2 = a^2/d.$$

从而当 $n \to +\infty$，$a^2 \sim nl^2$ 时

$$P(\xi_n^{(1)} \leqslant x_1, \cdots, \xi_n^{(d)} \leqslant x_d)$$

$$\sim \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} \frac{\exp\left\{-\sum_{i=1}^{d} dv_i^2/2a^2\right\}}{(\sqrt{2\pi/d}a)^d} \mathrm{d}v_1 \cdots \mathrm{d}v_d.$$

特别地，当 $d=2$，

$$P(\xi_n^{(1)} \leqslant x, \xi_n^{(2)} \leqslant y) \sim \int_{-\infty}^{x} \int_{-\infty}^{y} \frac{\exp(-(u^2+v^2)/a^2)}{\pi a^2} \mathrm{d}u\mathrm{d}v;$$

当 $d=3$，

$$P(\xi_n^{(1)} \leqslant x, \xi_n^{(2)} \leqslant y, \xi_n^{(3)} \leqslant z)$$

$$\sim \int_{-\infty}^{x} \int_{-\infty}^{y} \int_{-\infty}^{z} \frac{\exp\left\{-\frac{3}{2a^2}(u^2+v^2+w^2)\right\}}{(\sqrt{2\pi/3}a)^3} \mathrm{d}u\mathrm{d}v\mathrm{d}w;$$

**2. 自由链平均末端距**

$$E|\xi_n| \sim \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \sqrt{u^2+v^2+w^2} \frac{\exp\left\{-\frac{3}{2a^2}(u^2+v^2+w^2)\right\}}{(\sqrt{2\pi/3}a)^3} \mathrm{d}u\mathrm{d}v\mathrm{d}w$$

$$= \sqrt{\frac{8}{3\pi}} a \sim \sqrt{\frac{8}{3\pi}} (nl^2)^{1/2}.$$

**3. 自由链 Gauss 线团的形状($d=3$)**

为了表达自由链$\{\xi_k; k=1,2,\cdots,n\}$的形状，我们可以考虑它在若干个位置的联合分布. 例如考虑第 $i_1, i_2, \cdots, i_k$ 个键的末端的位置的联合分布密度可由以下正态分布密度近似($n \to +\infty$，$i_k$

$i_{k-1}\approx \alpha_k n$)

$$
\begin{aligned}
&p_{i_1,\cdots,i_r}(\underline{v}_1,\cdots,\underline{v}_r)\\
&\sim \prod_{k=1}^{r}\left\{\exp\left(-\frac{3|\underline{v}_k-\underline{v}_{k-1}|^2}{2(i_k-i_{k-1})l^2}\right)\Big/\left(\sqrt{\frac{2}{3}\pi(i_k-i_{k-1})}\,l\right)^3\right\}\quad (i_0=0)\\
&=\exp\left(-\sum_{k=1}^{r}\frac{3|\underline{v}_k-\underline{v}_{k-1}|^2}{2(i_k-i_{k-1})l^2}\right)\Big/\left(\prod_{k=1}^{r}\sqrt{\frac{2}{3}\pi(i_k-i_{k-1})}\right)^3 l^{3r}.
\end{aligned}
$$

由此我们还可进一步得到关于自由链卷曲成的“线团”的形状的一些特征数. 例如

1° 两端固定在$(0,0,0)$及$(0,0,h)$的自由链的第$j$(其中$j\approx \alpha n$)个链终点的位置$\xi_j\triangleq(x(\omega),y(\omega),z(\omega))$的分布密度

$$
s(x,y,z,h)=\frac{p_{0,j,n}(\underline{v}_1,\underline{v}_2,\underline{v}_3)}{p_{0,n}(\underline{v}_1,\underline{v}_3)},
$$

其中$\underline{v}_1=(0,0,0)$, $\underline{v}_3=(0,0,h)$.

$$
s(\xi_j=(x,y,z))=\frac{\exp\left[-\frac{3n}{2}\left(x^2+y^2+\left(z-\frac{jh}{n}\right)^2\right)\Big/l^2 j(n-j)\right]}{\left(\sqrt{\frac{2}{3n}\pi j(n-j)}\cdot l\right)^3}.
$$

由此容易求出该点坐标的均值为$\left(0,0,\frac{jh}{n}\right)$. 坐标的均方值为

$$
E(x(\omega)^2)=E(y(\omega)^2)=j(n-j)l^2/3n,
$$

$$
E(z(\omega)^2)=j(n-j)l^2/3n+(jh/n)^2,
$$

$$
E|\xi_j|^2=j(n-j)l^2/n+(jh/n)^2.
$$

这个结果首先由化学家 Katchalaky, Kuenzle 与 Kuhn 三人直观地得到(见 J. Polym. Sci, 5, P283(1950)).

2° 平均截面尺寸.

在高分子化学中称$(E|\xi_j|^2)^{1/2}=(nl^2)^{1/2}$为链的卷曲线团的均方根长度. 由于自由链的分布是各向同性的, 1°中的分布中可将$z$轴选作首末端的连线. 于是第$j$个链的末端到链的连接首末端的轴线的距离的均值:

$$
f(j)\triangleq(E(x(\omega)^2+y(\omega)^2))^{1/2}
$$

$$-(2j(n-j)l^2/3n)^{1/2} \quad (0 \leqslant j \leqslant n)$$

对 $j$ 在 $[0,n]$ 中的最大值是 $(nl^2/6)^{1/2}$，它可以看成此链的卷曲线团的"宽度".

3° 自由链在给定方向的平均最大投影.

有些化学家认为除了上述有关线团的平均参数外，需考虑一些最大尺寸的平均值. 例如对一个给定的方向 $\underline{e}$，问线团在这个方向上的投影(最大)长度是什么. 其实就是要求：

$$X \triangleq E(\max_{0\leqslant m\leqslant n} \underline{\xi}_m \cdot \underline{e} - \min_{0\leqslant m\leqslant n} \underline{\xi}_m \cdot \underline{e}).$$

如前，我们用 Brown 运动 $\{c\underline{B}_t; 0\leqslant t\leqslant 1\}$ 来近似 $\{\underline{\xi}_m; 0\leqslant m\leqslant n\}$ $(n \gg 1)$

$$X \approx E(c\max_{0\leqslant t\leqslant 1} \underline{B}_m \cdot \underline{e} - c\min_{0\leqslant t\leqslant 1} \underline{B}_m \cdot \underline{e}).$$

由于 $\underline{B}_s$ 在 $s$ 的正负方向是对称的，而且 $B_0=0$，所以 $\max\limits_{0\leqslant t\leqslant 1} \underline{B}_m \cdot \underline{e}$ 与 $-\min\limits_{0\leqslant t\leqslant 1} \underline{B}_m \cdot \underline{e}$ 同分布. 因此

$$X \approx 2cE(\max_{0\leqslant t\leqslant 1} \underline{B}_t \cdot \underline{e}).$$

另一方面 $\tilde{\boldsymbol{B}} \triangleq \{\tilde{B}_t \triangleq \underline{B}_t \cdot \underline{e}; t\geqslant 0\}$ 仍是一个一维标准 Brown 运动. 这是因为 $\tilde{B}_t$ 无非是 $\underline{B}_t$ 各分量的线性组合，所以 $\tilde{\boldsymbol{B}}$ 仍是独立增量的 Gauss 过程，而且

$$\begin{aligned}
&E|\tilde{B}_t - \tilde{B}_s|^2 \\
&\quad = E[(B_t^{(1)} - B_s^{(1)})^2 e_1^2 + (B_t^{(2)} - B_s^{(2)})^2 e_2^2 + (B_t^{(3)} - B_s^{(3)})^2 e_3^2] \\
&\quad = (t-s)E|\underline{e}|^2 = (t-s),
\end{aligned}$$

从而，按标准 Brown 运动的定义 $\tilde{\boldsymbol{B}}$ 是 1-维标准 Brown 运动. 进而我们由(2.22)得

$$E(\max_{0\leqslant t\leqslant 1} \tilde{B}_t) = EM_1 = \sqrt{\frac{2}{\pi}},$$

于是我们要计算的

$$X \approx 2cE(\max_{0\leqslant t\leqslant 1} \tilde{B}_t) = 2cEM_1 = 2c\sqrt{\frac{2}{\pi}}$$

$$\approx \sqrt{\frac{8}{\pi}}\sqrt{\frac{nl^2}{3}} = \sqrt{\frac{8}{3\pi}}(nl^2)^{1/2} \approx 0.9(nl^2)^{1/2}.$$

上式的意思是在给定方向的平均最大投影是均方根末端距的 0.9 倍，线团是约在一个长 $(nl^2)^{1/2}$ 宽 0.9 倍于长的范围中.

## §2.5 基本更新定理

### 1. 独立同分布列的 Wald 等式

**引理 2.3（Wald 等式）** 设 $\{\tau_k\}$ 为存在期望的独立同分布列，又设正整值随机变量 $N$ 与 $\{\tau_k\}$ 的发展一致，意即对 $\forall\ n$，$\{N \leqslant n\}$ 只依赖 $\tau_1,\cdots,\tau_n$，（从而与 $\tau_{n+1},\tau_{n+2},\cdots$ 独立）. 并且我们再设 $EN < \infty$. 那么

$$E\Big(\sum_{k=1}^{N}\tau_k\Big) = EN \cdot E\tau_1. \tag{2.23}$$

**证明** $\{N \geqslant k\} = 1 - \{N \leqslant k-1\}$，由假定应该与 $\tau_k,\tau_{k+1},\cdots$ 独立. 于是（记 $I_{\{\ \}}$ 为{ }的示性函数）

$$\begin{aligned}
E\Big(\sum_{k=1}^{N}\tau_k\Big) &= E\sum_{k=1}^{\infty}(\tau_k I_{\{N\geqslant k\}}) \\
&= \sum_{k=1}^{\infty}E(\tau_k I_{\{N\geqslant k\}}) \quad \text{（级数绝对收敛）} \\
&= \sum_{k=1}^{\infty}E\tau_k\, EI_{\{N\geqslant k\}} \quad \text{（独立性）} \\
&= \sum_{k=1}^{\infty}E\tau_1\, P(N \geqslant k) \quad \text{（同分布）} \\
&= E\tau_1\sum_{k=1}^{\infty}P(N \geqslant k) \\
&= E\tau_1 \cdot EN.
\end{aligned}$$

**定理 2（基本更新定理）** 设 $\{\tau_k\}$ 为正值随机变量列，又设 $N_t$

由下式定义

$$\sum_{k=1}^{N_t}\tau_k \leqslant t < \sum_{k=1}^{N_t+1}\tau_k, \tag{2.24}$$

那么

$$\frac{EN_t}{t} \to \frac{1}{E\tau_1}. \tag{2.25}$$

**证明** 首先注意 $N_t$ 满足引理 2.3 中关于 $N$ 的条件. 事实上

$$\{N_t \leqslant n\} = \bigcup_{l \leqslant n}\left\{\sum_{k=1}^{l}\tau_k \leqslant t < \sum_{k=1}^{l+1}\tau_k\right\},$$

于是 $\{N_t+1 \leqslant n\} = \{N_t \leqslant n-1\}$ 只依 $\tau_1,\cdots,\tau_n$, 从而与 $\tau_{n+1},\cdots$ 独立. 对(2.24)右半不等式当 $E\tau_1 < \infty$ 时, 用 Wold 等式得到

$$t \leqslant E\tau_1(EN_t+1),$$

因此当 $E\tau_1 < \infty$ 时有

$$\varliminf_{t\to\infty}\frac{EN_t}{t} \geqslant \frac{1}{E\tau_1}. \tag{2.26}$$

记

$$\tau_k^{(M)} = \begin{cases} \tau_k, & \text{当 } \tau_k \leqslant M; \\ M, & \text{当 } \tau_k \geqslant M. \end{cases}$$

那么 $E\tau_1^{(M)} < \infty$, 而且 $\{\tau_k^{(M)}\}$ 仍为独立同分布列. 记 $N_t^{(M)}$ 满足:

$$\sum_{k=1}^{N_t^{(M)}}\tau_k^{(M)} \leqslant t < \sum_{k=1}^{N_t^{(M)}+1}\tau_k^{(M)}. \tag{2.24$'$}$$

显见 $N_t \leqslant N_t^{(M)}$. 又由于 $\tau_k^{(M)} \leqslant M$ 及(2.24)$'$我们有

$$\sum_{k=1}^{N_t^{(M)}+1}\tau_k^{(M)} \leqslant t + M. \tag{2.27}$$

对此用 Wold 等式得:

$$E\tau_1^{(M)}(EN_t^{(M)}+1) \leqslant t + M.$$

从而

$$E\tau_1^{(M)}(EN_t+1) \leqslant t + M.$$

于是

$$\overline{\lim_{t\to\infty}}\frac{EN_t}{t}\leqslant\frac{1}{E\tau_1^{(M)}}.$$

令 $M\to\infty$，便得

$$\overline{\lim_{t\to\infty}}\frac{EN_t}{t}\leqslant\frac{1}{E\tau_1}. \tag{2.28}$$

特别地，当 $E\tau_1=\infty$ 时，由上式推出

$$\lim_{t\to\infty}\frac{EN_t}{t}=0=\frac{1}{E\tau}.$$

而当 $E\tau_1<\infty$ 时，由(2.28)与(2.26)一起推出

$$\lim_{t\to\infty}\frac{EN_t}{t}=\frac{1}{E\tau}.$$

## §2.6 附　　录

1. 有关 Brown 运动及随机徘徊的马氏性方面的问题，我们将在马氏过程与马氏链的有关章节中去阐述.

2.（DeMoivre 极限定理）：设 $\{x_k(\omega):k=1,2,\cdots\}$ 是相互独立同分布随机变量列，$P(x_k(\omega)=1)=p$，$P(x_k(\omega)=-1)=q=1-p$. 令 $S_n=\sum_{k=1}^{n}x_k$，则

$$\lim_{n\to\infty}P\left\{a<\frac{S_n-n(p-q)}{\sqrt{npq}}\leqslant b\right\}=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\mathrm{d}x.$$

此定理是定理 2.3 的特殊情况下的部分结果.

3. 定理 2.3 中 $x^{(n)}$ 与 $\tilde{x}^{(n)}$ 收敛到标准 Brown 运动的含义是弱收敛（或称按分布收敛）. 在此，我们给出其严格定义，进一步的内容可参见钱敏平、龚光鲁：《随机过程论》及 Billingsley：《Convergence of Probability Measures》.

设有度量空间 $\Omega$、$\mathscr{F}$ 是包含 $\Omega$ 的所有开子集的最小 $\sigma$ 代数. 在可测空间 $(\Omega,\mathscr{F})$ 上概率测度 $\{\mu_n:n=1,2,\cdots\}$ 与 $\mu$. 称 $\mu_n$ 弱收敛

到 $\mu$，如果对$\forall f\in C(\Omega)$有

$$\lim_{n\to\infty}\int f(\omega)\mu_n(\mathrm{d}\omega) = \int f(\omega)\mu(\mathrm{d}\omega).$$

其中 $C(\Omega)$是 $\Omega\to R$ 的全体有界连续泛函的集合：$C(\Omega)=\{f:\Omega\to R, f$ 有界连续$\}$，而 $f$ 连续的意义是$\forall\ \omega_n,(n\geqslant 1),\omega\in\Omega$，由$\lim\limits_{n\to\infty}\omega_n=\omega$ 有$\lim\limits_{n\to\infty}f(\omega_n)=f(\omega)$.

例如

$$\Omega = C[0,+\infty) = \{f:[0,+\infty)\to R\},$$

令

$$\rho_n(f,g) = \sup_{x\in[0,n)}|f(x)-g(x)|,$$

$$\rho \triangleq \sum_{n=1}^{\infty}\frac{\rho_n}{2^n}$$

是一个度量空间. $\omega_n\in\Omega$，令 $\omega_n\triangleq\{\omega_n(t);t\geqslant 0\}$，则 $\omega_n\to\omega\in\Omega(n\to+\infty)$当且仅当连续函数 $\omega_n$ 在任意闭区间上一致收敛到连续函数 $\omega$. 容易看出对于 $n$ 维连续函数 $\varphi$，当 $f(\omega)\triangleq\varphi(\omega(t_1),\omega(t_2),\cdots,\omega(t_n))(n\geqslant 1)$时，$f(\omega)$一定是一个 $\Omega$ 上的连续泛函. $\{\tilde{x}^{(n)}\}$的收敛可看成在 $\Omega=C[0,+\infty)$上的弱收敛.

又例如 $\Omega=D(0,+\infty)=\{\omega=(\omega(t);t\geqslant 0);\omega(\cdot)$是右连续有左极限的函数$\}$，可在其上定义距离(Skorohod 距离)而成为一个度量空间. 在 §2.3 中 $x_t^{(n)}$ 就是定义在这个空间上的，因为 $x_\cdot^{(n)}(\omega)$ 的可能轨道是阶梯函数，因而$\{x^{(n)}\}$的收敛可看成 $\Omega=D(0,+\infty)$ 上的弱收敛.

4. 本章 §2.4 中自由连接高分子链由著名物理化学家 P. J. Flory 首创，是高分子物理化学中基本而影响重大的模型. 由此，还有一系列更复杂更接近实际高分子的模型及更深入的结论. 本章内容参考吴大诚：《高分子构象统计理论导引》一书.

5. 由定义 2.1，Brown 运动的轨道应该是连续的. 但若按 Kolmogorov 定理构造遵从 Brown 分布(即满足 $B.1$)与 $B.2$))的随机过程 $\boldsymbol{B}$，并不能立即自然得到对几乎所有的轨道 $\omega$，$B.(\omega)$是

连续函数. 但可以证明这个结论是正确的. 不仅如此，还可以证明对几乎所有的轨道，$B_\cdot(\omega)$对每个时间 $t$ 对时间的导数都不存在（这正符合 Brown 运动的物理直观：粒子每时每刻都在改变运动方向）. 关于 Brown 运动的轨道性质还有一系列更深刻的结果，读者可参见钱敏平，龚光鲁：《随机过程论》及 D. Revuz 与 M. Yor：《Continuous Martingales and Brownian Motion》.

## 习　　题

1. 在简单对称随机徘徊中，设 $L_n$ 为从第 $n$ 次开始连续向右走的次数，求 $L_n$ 的分布.

2. 设钱币出国徽的概率均为 $p$，每隔单位时间独立地扔两次，设在第 $n$ 单位时间

$$\begin{cases} X_n = \begin{cases} 1, & \text{先出国徽，后出另一面；} \\ 0, & \text{两次结果一致；} \\ -1, & \text{其他；} \end{cases} \\ \tau = T_1 \wedge T_{-1} \end{cases}$$

（$T_y$ 定义请见第 12 页及(2.4)，其中 $\xi_0(\omega)=0$）. 求 $P(\tau<\infty)$，计算 $E\tau$ 及 $X_\tau$ 的分布.

3. 计算 $\mathrm{cov}(\xi_n,\xi_m)$，其中 $\xi_n$ 为随机徘徊.

4. 设 $Z$ 上简单随机徘徊首次返回 0 的时刻为 $\eta$，求证

$$P(\eta < \infty) = pP_1(T_0 < \infty) + qP_{-1}(T_0 < \infty) = 2(p \wedge q)$$

（其中 $P_1,P_{-1}$ 分别指从 1，$-1$ 出发，$T_0$ 是过程首达 0 的时刻）.

5. 设 $Z$ 上对称简单随机徘徊为 $\xi_n$，求证 $u(n,x)\triangleq P_m(\xi_n=x)$（$m$ 表示出发点）满足

$$\begin{cases} u(n+1,x) - u(n,x) \\ \qquad = \dfrac{1}{2}(u(n,x+1) + u(n,x-1) - 2u(n,x)), \\ u(0,x) = \delta_{xm}. \end{cases}$$

6. 设直线上两质点初始位置分别在 $x$，$y$，每次按等概率任选一点，该点作一次简单随机徘徊，求两点最后相遇的概率(注：考虑两点距离).

7. 设 $Z$ 上有两个相互独立的随机徘徊 $\{\xi_n\}$，$\{\eta_n\}$，又设 $\xi_n - \xi_{n-1}$ 与 $\eta_n - \eta_{n-1}$ 有同分布 $\{p_x : x \in Z\}$，求证 $\zeta_n \triangleq \xi_n - \eta_n$ 的分布是对称的，再求 $\zeta_n - \zeta_{n-1}$ 的分布.

8. 设 $\{\xi_n\}$ 为 $Z$ 上简单随机徘徊，$\xi_0 = 0$，求证

$$\sum_{n=0}^{\infty} P(\xi_n = 0) = \begin{cases} +\infty, & p = q; \\ (1-4pq)^{-1/2} = |p-q|^{-1}, & p \neq q. \end{cases}$$

9. 求 $E\{\xi_n I_{\{\max\limits_{1 \le m \le n} \xi_m < a\}} + a I_{\{\max\limits_{1 \le m \le n} \xi_m \ge a\}}\}$.

10. 设直线上 Brown 运动首达 $x$ 的时刻为 $\tau_x$，又设 $B_0 = 0$，求 $E\tau_x$，$Ee^{-\lambda\tau_x}$. 进一步证明对 $\forall\ y$ 有

$$P\left\{\frac{1}{n}\tau_{\sqrt{n}} \leqslant y\right\} \xrightarrow{n \to \infty} P(\tau_1 \leqslant y).$$

11. 证明 Brown 运动 $B_t$ 有

$$P(B_t \text{ 在}(s,t)\text{ 有零点} \mid B_s = x) = P_0(\tau_x \leqslant t - s).$$

12. 证明 $R$ 上对称简单随机徘徊 $\{\xi_n\}$，$\xi_0 = 0$，有

$$P\{\max_{m \le n} \xi_m \geqslant y\} = 2P(\xi_n \geqslant y) - P(\xi_n = y).$$

叙述 Brown 运动的对应定理.

13. 在破产问题中，求赌博停止时(输光或赢得 $b - x$ 元)，赌徒平均净赢钱数.

14. 假设已知

$$\phi_x = E_x(T_0(\omega) \wedge T_b(\omega)) < \infty.$$

求证它满足

$$\begin{cases} \phi_x = p\phi_{x+1} + q\phi_{x-1} + 1 & (0 < x < b), \\ \phi_0 = \phi_b = 0. \end{cases}$$

试求出 $\phi_x$.

15. 设

$$u_{x,n} = P_x(T_0(\omega) = n) \quad (\text{破产时间分布}),$$

求证它的矩母函数

$$u_x(z) \triangleq \sum_{n=0}^{\infty} u_{x,n} z^n$$

满足

$$\begin{cases} u_x(z) = pzu_{x+1}(z) + qzu_{x-1}(z), \\ u_0(z) = 1, u_b(z) = 0. \end{cases}$$

这是(把 $z$ 看成参数)一个 $x$ 的常系数二阶差分方程的边值问题. 设差分方程的基本解为 $\lambda_1^x, \lambda_2^x$. 于是通解为

$$c_1(z)(\lambda_1(z))^x + c_2(z)(\lambda_2(z))^x.$$

试具体求出 $u_x(z)$, ($\lambda_1(z)$, $\lambda_2(z)$)并用边值条件确定 $c_1(z)$, $c_2(z)$.

16. 设赌徒每次赌博可能有赢 1 元、输 1 元或不赢不输三种结果, 概率各为 $p, q, r$ ($p+q+r=1$), 试把破产问题及上面 1 至 3 推广到这一情形.

17. 设一切同赌徒破产问题, 只改变一点: 当赌徒只有 1 元时, 以概率 $p$ 赢 1 元, 以概率 $(1-\delta)q$ 输 1 元, 以概率 $\delta q$ 不输不赢, 求破产概率满足的方程与边值条件(称为弹性边界的随机徘徊).

18. 平面上从原点引三条射线, 夹角各为 $\frac{2\pi}{3}$, 在各条射线上的整点分别记成 $1, 2, 3, \cdots; 1', 2', 3', \cdots; 1'', 2'', 3'', \cdots$ 称为整点. 试在这整点上定义对称随机徘徊, 并求开始时质点在 $m$ 经过 $n$ 个单位时间后, 质点分别在 $m'$ 与 $l$ 的概率:

$$P_m(x_n = m'), \qquad P_m(x_n = l).$$

19. 甲乙两个候选人各得选票 $n$ 张与 $m$ 张($n>m$), 设开票过程中甲始终先于乙的概率记为 $p_{n,m}$, 试导出 $p_{n,m}$ 应满足的差分方程, 然后证明

$$p_{n,m} = \frac{n-m}{n+m}.$$

20. 设 $\xi_n$ 是对称随机徘徊, 且 $\xi_0 = 0$.

(1) 令 $u_n \triangleq P(\xi_{2n}=0)$，证明 $u_n=\frac{(2n-1)!!}{(2n)!!}$；

(2) 证明 $\sum_{k=1}^{n}\frac{u_k}{2k-1}+u_n=1$；

(3) 证明 $\xi$ 首次返回 0 的时刻为 $2n$ 的概率

$$P(\xi_1\neq 0,\cdots,\xi_{2n-1}\neq 0,\xi_{2n}=0)=\frac{u_n}{2n-1};$$

(4) 用(2)，(3)证明 $\xi$ 在 $2n$ 前从未返回 0 的概率

$$P(\xi_1\neq 0,\cdots,\xi_{2n}\neq 0)=P(\xi_{2n}=0);$$

(5) 证明：到 $2n$ 时刻为止，$\xi_m$ 最后一次到达 0 的时刻 $\tau$ 的分布为

$$P(\tau=2k)=u_k u_{n-k}\quad(k<n);$$

(6) 证明 $k$ 很大时

$$u_k u_{n-k}\sim\frac{1}{\pi}\frac{1}{\sqrt{k(n-k)}}.$$

注：(5)中的分布 $u_k u_{n-k}$ 称为离散的反正弦分布，这是因为对于 $\forall\, x\in(0,1)$ 粗略地

$$\begin{aligned}
P\left(\frac{\tau}{2n}\leqslant x\right)&=\sum_{k=0}^{nx}P(\tau=2k)=\sum_{k=0}^{nx}u_k u_{n-k}\\
&=\sum_{\frac{k}{n}=0}^{x}u_k u_{n-k}=\lim_{\varepsilon\to 0}\sum_{\varepsilon\leqslant\frac{k}{n}\leqslant x}u_k u_{n-k}\\
&\overset{(*)n\text{大}}{\approx}\lim_{\varepsilon\to 0}\sum_{\varepsilon\leqslant\frac{k}{n}\leqslant x}\frac{1}{\pi k}\frac{1}{\sqrt{\frac{k}{n}\left(1-\frac{k}{n}\right)}}\\
&\approx\lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{\varepsilon}^{x}\frac{\mathrm{d}y}{\sqrt{y(1-y)}}=\frac{1}{\pi}\int_{0}^{x}\frac{\mathrm{d}y}{\sqrt{y(1-y)}}\\
&=\frac{2}{\pi}\arcsin\sqrt{x},
\end{aligned}$$

这里(*)处用了两种近似“$n\to\infty$”与 $\varepsilon\to 0$ 的交换性，它的合法性我们暂且不管它(需用测度论的知识)．于是我们有当 $n$ 很大时，$\frac{\tau}{2n}$ 近似分布为反正弦律

$$P(\xi_t \neq 0, \forall\ 2nx < t \leqslant 2n) \overset{n大}{\approx} \frac{2}{\pi}\arcsin \sqrt{x}\ .$$

(7) 设 $T_0 = \inf\{n \geqslant 1 : \xi_n = 0\}$，试利用(4)及(6)证明 $ET_0 = \infty$。

# 第三章　离散时间参数 Markov 链(马氏链)

考虑比简单随机徘徊较为一般的模型：一个粒子在 1 维格点集 Z 上运动，每步随机地走一格．当粒子在位置 $x(\in Z)$时，它向左或向右走一格的概率分别是 $p_x$ 与 $q_x=1-p_x$．现在我们考虑的这一随机运动与简单随机徘徊的基本差别是：后者是独立增量的(即每步位移是相互独立的)；而前者不然，每步位移依赖于它现在的位置 $x$——此步之前的各步位移的总和．然而，在现在的位置 $x$ 给定的条件下粒子以后的运动却仍然与粒子在过去的运动独立．这种统计特点就叫 Markov 性(这个名字由它的首创人 Markov 而得名，因而简称马氏性)．具有 Markov 性的随机过程称为 Markov 过程，简称马氏过程．马氏过程是一大类广泛适用于各种领域的随机过程．特别是参数离散、状态空间有限或可数的情况(称离散参数马氏链)，由于它的模型简洁但又抓住事物的本质，它的性质也为人们认识得特别清楚，因而是一类特别重要的特殊随机过程．本章主要讨论它的基本概念，决定统计规律(分布)的特征——转移概率矩阵以及统计性质．

## §3.1　Markov 链的概念与转移阵

### 1．概念

**定义 3.1(马氏链)**　一个随机过程 $\xi\triangleq\{\xi_0,\xi_1,\cdots,\xi_n,\cdots\}$ 称为一个离散参数的 Markov 链，如果它的状态空间有限或可数，并且对任意的 $m\geqslant1,n\geqslant0$ 及状态 $i,j,i_0,i_1,\cdots,i_{n-1}$ 都有

$$P(\xi_{n+m}=j\mid\xi_n=i,\xi_{n-1}=i_{n-1},\cdots,\xi_0=i_0)$$

$$= P(\xi_{n+m} = j | \xi_n = i). \tag{3.1}$$

记

$$p_{ij}(n,m) = P(\xi_m = j | \xi_n = i) \quad (m \geqslant n). \tag{3.2}$$

由于状态空间是有限或可数的，不妨设它是$\{1,2,\cdots,N\}$(状态空间元素个数是 $N$)或是$\{1,2,\cdots,N,\cdots\}$或 $Z$(状态空间可数)，我们可将全部 $p_{ij}(n,m)$ 排成矩阵：$\boldsymbol{P}(n,m)=(p_{ij}(n,m))$，于是 $\boldsymbol{P}(n,m)$具有以下性质：

**命题 3.1** $P.1)$ 对任意的状态 $i$，$\sum_j p_{ij}(n,m) = 1$，即 $\boldsymbol{P}(n,m)\boldsymbol{1}=\boldsymbol{1}$，其中 $\boldsymbol{1}$ 为元素全为 1 的列向量，其维数与状态空间的元素个数一致.

$P.2)$ $0 \leqslant p_{ij}(n,m) \leqslant 1$；$p_{ij}(n,n)=\delta_{ij}$ （$\delta_{ij}$为 0，如 $i\neq j$；为 1，如 $i=j$).

$P.3)$ 对任意的 $0\leqslant n\leqslant m\leqslant l$，$\sum_k p_{ik}(n,m)p_{kj}(m,l) = p_{ij}(n,l)$，此即

$$\boldsymbol{P}(n,m)\boldsymbol{P}(m,l) = \boldsymbol{P}(n,l).$$

**证明** $P.2)$显然. 又

$$\begin{aligned}\sum_j p_{ij}(n,m) &= \sum_j P(\xi_m = j | \xi_n = i) \\ &= P\{\bigcup_j \{\xi_m = j\} | \xi_n = i\} \\ &= P(\Omega | \xi_n = i) = 1,\end{aligned}$$

$$\begin{aligned}\sum_k p_{ik}(n,m)p_{kj}(m,l) &= \sum_k P(\xi_m = k | \xi_n = i)P(\xi_l = j | \xi_m = k) \\ &= \sum_k P(\xi_m = k | \xi_n = i)P(\xi_l = j | \xi_m = k, \xi_n = i) \\ &= \sum_k P(\xi_m = k, \xi_l = j | \xi_n = i) \\ &= P\{\bigcup_k \{\xi_m = k, \xi_l = j\} | \xi_n = i\} \\ &= P(\xi_l = j | \xi_n = i).\end{aligned}$$

**定义 3.2** 满足条件 $P.1)$，$P.2)$，$P.3)$的矩阵族$\{\boldsymbol{P}(n,m)\}$称

为一族转移概率阵. 满足 $P.1$)与 $P.2$)的一个矩阵称为一个随机矩阵.

**命题 3.2** 以下各命题彼此等价:

1) $\xi$ 是一个马氏链;

2) (3.1)式对 $m=1, n\geqslant 0$ 及状态 $i, j, i_0, \cdots, i_{n-1}$ 成立;

3) 对任意的 $0\leqslant n_1<n_2<\cdots<n_s$ 及状态 $i_1, i_2, \cdots, i_s$,

$$
\begin{aligned}
&P(\xi_{n_s}=i_s, \xi_{n_{s-1}}=i_{s-1}, \cdots, \xi_{n_1}=i_1) \\
&\quad = P(\xi_{n_1}=i_1)p_{i_1i_2}(n_1, n_2)p_{i_2i_3}(n_2, n_3)\cdots p_{i_{s-1}i_s}(n_{s-1}, n_s);
\end{aligned} \tag{3.3}
$$

4) 在 3)中对于 $n_1=0, n_2=1, \cdots, n_s=s-1$ 及状态 $i_0, i_1, \cdots, i_{s-1}$ (3.3)式均成立;

5) 对于任意的 $m_1, \cdots, m_s, n_1, \cdots, n_r\geqslant 1, n>n_k(k\geqslant 1)$ 及状态 $j_1, \cdots, j_s, i, i_0, \cdots, i_r$ 均有

$$
\begin{aligned}
&P(\xi_{n+m_s}=j_s, \cdots, \xi_{n+m_1}=j_1 | \xi_n=i; \xi_{n-n_1}=i_1, \cdots, \xi_{n-n_r}=i_r) \\
&\quad = P(\xi_{n+m_s}=j_s, \cdots, \xi_{n+m_1}=j_1 | \xi_n=i);
\end{aligned} \tag{3.4}
$$

6) 在 5)中将(3.4)改为

$$
\begin{aligned}
&P(\xi_{n+m_s}=j_s, \cdots, \xi_{n+m_1}=j_1; \xi_{n-n_1}=i_1, \cdots, \xi_{n-n_r}=i_r | \xi_n=i) \\
&\quad = P(\xi_{n+m_s}=j_s, \cdots, \xi_{n+m_1}=j_1 | \xi_n=i) \\
&\quad\quad \times P(\xi_{n-n_1}=i_1, \cdots, \xi_{n-n_r}=i_r | \xi_n=i).
\end{aligned}
$$

此即对任意的状态 $i$,已知 $\xi_n=i$ 的条件下,过去与将来独立. $\xi_n=i$ 就是说,将"现在"看作时刻 $n$ 时,"现在"$\xi$ 所处的状态已知.

**证明** 显然有 5)$\Leftrightarrow$6), 5)$\Rightarrow$1)$\Rightarrow$2), 3)$\Rightarrow$4). 因而只需证明 2)$\Rightarrow$4)$\Rightarrow$3)及 3)$\Rightarrow$5). 从 2)由概率的乘法公式我们可得到 4):

$$
\begin{aligned}
&P(\xi_{n+m}=i_m, \xi_{n+m-1}=i_{m-1}, \cdots, \xi_n=i_0) \\
&\quad = P(\xi_n=i_0)P(\xi_{n+1}=i_1 | \xi_n=i_0)\cdots \\
&\quad\quad \times P(\xi_{n+m}=i_m | \xi_{n+m-1}=i_{m-1}, \cdots, \xi_n=i_0) \\
&\quad = P(\xi_n=i_0)p_{i_0i_1}(n, n+1)p_{i_1i_2}(n+1, n+2)\cdots
\end{aligned}
$$

$$\times p_{i_{m-1}i_m}(n+m-1,n+m).$$

又如在上式中令 $m=n_s$，并对全部可能的 $i_k(k\neq n_1,n_2,\cdots,n_s)$ 求和，左边就是

$$P(\xi_{n_s}=i_s,\xi_{n_{s-1}}=i_{s-1},\cdots,\xi_{n_1}=i_1).$$

而由 $P.1)$ 与 $P.2)$ 得到右边相应的和是

$$P(\xi_{n_1}=i_1)p_{i_1i_2}(n_1,n_2)\cdots p_{i_{s-1}i_s}(n_{s-1},n_s).$$

于是就得到 4)⇒3). 而由条件概率的定义与 3)，我们有

$$P(\xi_{n+m_s}=j_s,\cdots,\xi_{n+m_1}=j_1|\xi_n=i,\xi_{n-n_1}=i_1,\cdots,\xi_{n-n_r}=i_r)$$

$$=\frac{P(\xi_{n+m_s}=j_s,\cdots,\xi_{n+m_1}=j_1;\xi_n=i,\xi_{n-n_1}=i_1,\cdots,\xi_{n-n_r}=i_r)}{P(\xi_n=i,\xi_{n-n_1}=i_1,\cdots,\xi_{n-n_r}=i_r)}$$

$$=\frac{P(\xi_{n-n_r}=i_r)p_{i_ri_{r-1}}(n-n_r,n-n_{r-1})\cdots p_{i_1i}(n-n_1,n)}{P(\xi_{n-n_r}=i_r)}$$

$$\times\frac{p_{ij_1}(n,n+m_1)\cdots p_{j_{s-1}j_s}(n+m_{s-1},n+m_s)}{p_{i_ri_{r-1}}(n-n_r,n-n_{r-1})\cdots p_{i_1i}(n-n_1,n)}$$

$$=\frac{P(\xi_n=i)p_{ij_1}(n,n+m_1)\cdots p_{j_{s-1}j_s}(n+m_{s-1},n+m_s)}{P(\xi_n=i)}$$

$$=P(\xi_{n+m_1}=j_1,\cdots,\xi_{n+m_s}=j_s|\xi_n=i).$$

此即 3)⇒5). 命题证毕.

上面的命题说明： Markov 性就是"在已知现在的条件下，过去与将来独立".

## 2. 时齐马氏链

**定义 3.3(时齐性)** 离散时间参数马氏链 $\xi$ 称为时齐的，如果：

$$P(\xi_{n+m}=j|\xi_m=i)=P(\xi_n=j|\xi_0=i)\triangleq p_{ij}(n),$$

即它与 $m$ 无关. 这时矩阵 $\boldsymbol{P}(n)\triangleq(p_{ij}(n))$ 称为 $\xi$ 的 $n$ 步转移阵，特别将 $\boldsymbol{P}(1)$ 记为 $\boldsymbol{P}$，$p_{ij}\triangleq p_{ij}(1)$，称 $\boldsymbol{P}$ 为 $\xi$ 的转移阵.

一个时齐马氏链的转移阵 $\boldsymbol{P}$ 之所以特别重要，是因为此时在第 1 节中的

$$P(n,m) = P(n,n+1)\cdots P(m-1,m) = P^{m-n}.$$

可见 $P$ 决定了全部的 $P(\xi_m = j \mid \xi_n = i)$，从而由命题 3.2 的 3) 与 4)，$P$ 和初分布 $\mu = \{\mu_i\}$，($\mu_i = P(\xi_0 = i)$)就决定了 $\xi$ 的一切有限个时刻的联合分布，也就决定了 $\xi$ 的分布：

$$P(\xi_n = j) = \sum_i \mu_i p_{ij}(n),$$

其中 $p_{ij}(n) = (P^n)_{ij}$；而

$$\begin{aligned} & P(\xi_{n_1} = j_1, \cdots, \xi_{n_k} = j_k) \\ & \quad = \sum_i \mu_i p_{ij_1}(n_1) p_{j_1 j_2}(n_2 - n_1) \cdots p_{j_{k-1} j_k}(n_k - n_{k-1}). \end{aligned}$$

特别，令 $\mu_j(n) = P(\xi_n(\omega) = j)$，就有

$$\begin{aligned} \mu_j(n+1) &= \sum_k P(\xi_n(\omega) = k, \xi_{n+1}(\omega) = j) \\ &= \sum_k \mu_k(n) p_{kj} \quad (\textbf{Master 方程或主方程}). \end{aligned}$$

### 3. 几个简单的应用例题

**例 3.1** 广告效益的推算.

某甲种香波改变了广告方式，经调查发现：买甲香波及另外三种主要竞争对手的顾客每两个月的平均转换率如下表：

| 双月平均转换率（转换前＼转换后） | 甲 | 乙 | 丙 | 丁 |
|---|---|---|---|---|
| 甲 | 0.95 | 0.02 | 0.02 | 0.01 |
| 乙 | 0.30 | 0.60 | 0.06 | 0.04 |
| 丙 | 0.20 | 0.10 | 0.70 | 0 |
| 丁 | 0.20 | 0.20 | 0.10 | 0.50 |

假设目前的顾客分布率为(0.25，0.30，0.35，0.10)，求半年后甲香

波的顾客率.

**解** 令 $\boldsymbol{P}$ 为上表所列的转移阵;半年后,顾客在四种香波上的分布率是

$$v \triangleq (v_1, v_2, v_3, v_4) = (\mu_1, \mu_2, \mu_3, \mu_4)\boldsymbol{P}^3. \tag{3.1}$$

其中$(\mu_1, \mu_2, \mu_3, \mu_4) = (0.25, 0.30, 0.35, 0.10)$,将前面表中的具体数字代入,得到:

$$\boldsymbol{P}^2 = \begin{pmatrix} 0.9145 & 0.035 & 0.0352 & 0.0153 \\ 0.485 & 0.38 & 0.088 & 0.047 \\ 0.36 & 0.134 & 0.5 & 0.006 \\ 0.37 & 0.231 & 0.136 & 0.26 \end{pmatrix},$$

$$\boldsymbol{P}^3 = \begin{pmatrix} 0.8894 & & & \\ 0.60175 & & * & \\ 0.4834 & & & \\ 0.5009 & & & \end{pmatrix},$$

$$v_1 = (0.25 \quad 0.3 \quad 0.35 \quad 0.1)\begin{pmatrix} 0.8894 \\ 0.60175 \\ 0.4834 \\ 0.5009 \end{pmatrix} \doteq 0.624.$$

可见,甲种香波的顾客半年后将由原来的25%增至62%,新的广告方式很见效益.

**例 3.2** 简单随机徘徊作为 Markov 链.

在第二章中考虑的简单随机徘徊$\xi = \{\xi_n, n \geqslant 0\}$(其中

$$\xi_n(\omega) = \sum_{k=1}^{n} X_k(\omega) + X_0(\omega),$$

而$\{X_k; k \geqslant 0\}$为i.i.d.序列)是一个马氏链,这是因为

$$\begin{aligned} P(\xi_{n+1} = j \mid \xi_n = i, \xi_{n-1} = i_{n-1}, \cdots, \xi_0 = i_0) \\ = P(X_{n+1} = j - \xi_n \mid \xi_n = i, \xi_{n-1} = i_{n-1}, \cdots, \xi_0 = i_0) \\ = P(X_{n+1} = j - i \mid \xi_n = i, \xi_{n-1} = i_{n-1}, \cdots, \xi_0 = i_0) \\ = P(X_{n+1} = j - i) \quad (\text{因 } \xi_n, \cdots, \xi_0 \text{ 与 } X_{n+1} \text{ 独立}) \end{aligned}$$

$$
= \begin{cases} p, & j = i + 1; \\ q, & j = i - 1; \\ 0, & \text{其他} \end{cases}
$$

$$
= P(\xi_{n+1} = j \mid \xi_n = i).
$$

这时，$\xi = \{\xi_n, n \geqslant 0\}$是时齐的马氏链，它的转移概率阵是

$$
P = (p_{ij}) = \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & q & 0 & p & 0 & 0 & \cdots \\ \cdots & 0 & q & 0 & p & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & 0 & q & 0 & p & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.
$$

**例 3.3** 带吸收壁的随机徘徊.

设在 $a,b$ 两处分别立一吸收壁（见§2.2.4 段）. 令 $T_a(\omega)$，$T_b(\omega)$ 与 $\tau(\omega)$ 分别为随机徘徊首达 $a,b$ 及 $a$ 或 $b$ 之一的时间，因而 $\tau = T_a(\omega) \wedge T_b(\omega)$. 又令

$$
\eta_n(\omega) \triangleq \xi_{n\wedge\tau}(\omega), \qquad \eta \triangleq \{\eta_n; n \geqslant 0\},
$$

$$
\bar{\eta}_n(\omega) \triangleq \xi_{n\wedge T_a(\omega)}(\omega); \qquad \bar{\eta} \triangleq \{\bar{\eta}_n; n \geqslant 0\}.
$$

则 $\eta, \bar{\eta}$ 分别是在 $a$ 与 $b$ 两处都有吸收壁和仅 $a$ 一处有吸收壁的随机徘徊，它们都是马氏链.

下面我们来证明 $\eta$ 是马氏链，而把关于 $\bar{\eta}$ 是马氏链的证明留给读者. 由于

$$
P(\eta_{n+1} = j \mid \eta_n = i, \eta_{n-1} = i_{n-1}, \cdots, \eta_0 = i_0)
$$

$$
= \frac{P(\eta_0 = i_0, \cdots, \eta_{n-1} = i_{n-1}, \eta_n = i, \eta_{n+1} = j)}{P(\eta_0 = i_0, \cdots, \eta_{n-1} = i_{n-1}, \eta_n = i)}
$$

$$
= \frac{P(\xi_{0\wedge\tau} = i_0, \cdots, \xi_{(n-1)\wedge\tau} = i_{n-1}, \xi_{n\wedge\tau} = i, \xi_{(n+1)\wedge\tau} = j)}{P(\xi_{0\wedge\tau} = i_0, \cdots, \xi_{(n-1)\wedge\tau} = i_{n-1}, \xi_{n\wedge\tau} = i)}.
$$

又注意到 $\tau > n$ 当且仅当 $\xi_0, \xi_1, \cdots, \xi_n$ 都不是 $a$ 或 $b$，而当 $i_0, \cdots, i_{n-1}, i \neq a, b$ 时，上式等于

$$= \frac{P(\xi_0 = i_0, \cdots, \xi_{n-1} = i_{n-1}, \xi_n = i, \xi_{n+1} = j)}{P(\xi_0 = i_0, \cdots, \xi_{n-1} = i_{n-1}, \xi_n = i)}$$

$$= \frac{P(\xi_0 = i_0) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i} p_{ij}}{P(\xi_0 = i_0) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i}} = p_{ij}. \qquad (3.5\text{-}1)$$

当 $\tau \leqslant n$，在 $m < \tau$ 时 $\xi_m \neq a$ 或 $b$，而在 $m \geqslant \tau$ 时 $\xi_m = \xi_\tau$，因而这时只要考虑 $\tau = k, i_0, \cdots, i_{k-1} \neq a, b, i_k = \cdots = i_{n-1} = i = a$ 或 $b$ 的情形：

$$P(\eta_0 = i_0, \cdots, \eta_{n-1} = i_{n-1}, \eta_n = i, \eta_{n+1} = j)$$

$$= P(\eta_0 = i_0, \cdots, \eta_{n-1} = i_{n-1}, \eta_n = i, \eta_{n+1} = j, \tau = k)$$

$$= P(\xi_0 = i_0, \cdots, \xi_{k-1} = i_{k-1}, \xi_k = \cdots = \xi_{n-1} = \xi_n = i, \xi_{(n+1)\wedge k} = j)$$

$$= P(\xi_0 = i_0, \cdots, \xi_{k-1} = i_{k-1}, \xi_k = \cdots = \xi_{n-1} = \xi_n = i, \xi_k = j)$$

$$= P(\xi_0 = i_0) p_{i_0 i_1} \cdots p_{i_{n-1} i} \delta_{ij}$$

$$= P(\xi_0 = i_0, \xi_1 = i_1, \cdots, \xi_{n-1} = i_{n-1}, \xi_n = i) \delta_{ij}$$

$$= P(\xi_0 = i_0, \xi_1 = i_1, \cdots, \xi_{n-1} = i_{n-1}, \xi_n = i, \tau = k) \delta_{ij}$$

$$= P(\eta_0 = i_0, \cdots, \eta_{n-1} = i_{n-1}, \eta_n = i) \delta_{ij}.$$

于是当 $i_0, \cdots, i_{k-1} \neq a, b, i_k = \cdots = i_{n-1} = i = a$ 或 $b$ 时，有

$$P(\eta_{n+1} = j | \eta_n = i, \eta_{n-1} = i_{n-1}, \cdots, \eta_0 = i_0)$$

$$= \delta_{ij} = P(\eta_{n+1} = j | \eta_n = i). \qquad (3.5\text{-}2)$$

综合(3.5-1)与(3.5-2)

$$P(\eta_{n+1} = j | \eta_n = i, \eta_{n-1} = i_{n-1}, \cdots, \eta_0 = i_0)$$

$$= \begin{cases} \delta_{ij}, & \exists\, k \in [0, n], \text{使 } i_0, \cdots, i_{k-1} \neq a, b, i_k = \cdots = i_{n-1} = i = a \text{ 或 } b; \\ p_{ij}, & \text{其他情形}. \end{cases} \qquad (3.6)$$

另一方面，当 $i \neq a$ 或 $b$ 时，我们有

$$P(\eta_n = i, \eta_{n+1} = j) = P(\eta_n = i, \tau > n, \eta_{n+1} = j).$$

但是，当 $\tau > n$ 时，$\tau \geqslant n+1$，因而 $\eta_{n+1} = \xi_{(n+1)\wedge\tau} = \xi_{n+1}$，

$$\text{上式} = P(\xi_n = i, \tau > n, \xi_{n+1} = j)$$

$$= P(\xi_{n+1} = j | \xi_n = i, \tau > n) P(\xi_n = i, \tau > n)$$

$$= P(\xi_{n+1} = j | \xi_n = i, \xi_0, \cdots, \xi_{n-1} \neq a, b) P(\eta_n = i).$$

由 $\xi$ 的马氏性，上式等于

$$P(\xi_{n+1}=j|\xi_n=i)P(\eta_n=i)=P(\eta_n=i)p_{ij}.$$

可见这时有

$$P(\eta_{n+1}=j|\eta_n=i)=p_{ij}. \tag{3.7}$$

当 $i=a$ 或 $b$ 时，

$$\begin{aligned}P(\eta_n=i,\eta_{n+1}=j)&=P(\eta_n=i,\eta_{n+1}=j,\tau\leqslant n)\\&=P(\xi_n=i,\tau\leqslant n;\xi_{(n+1)\wedge\tau}=j)\\&=\delta_{ij}P(\xi_n=i,\tau\leqslant n)\\&=P(\eta_n=i)\delta_{ij}.\end{aligned}$$

于是当 $i=a$ 或 $b$ 时，

$$P(\eta_{n+1}=j|\eta_{n+1}=i)=\delta_{ij}. \tag{3.8}$$

综合(3.6)至(3.8)就得到 $\eta$ 的马氏性，而且

$$P(\eta_{n+1}=j|\eta_n=i,\eta_{n-1}=i_{n-1},\cdots,\eta_0=i_0)=\begin{cases}p_{ij}, & \text{当 } i\neq a,b;\\ \delta_{ij}, & \text{当 } i=a \text{ 或 } b.\end{cases}$$

于是，$\eta$ 的转移概率阵就是

$$\boldsymbol{P}=(p_{ij})=\begin{pmatrix}1 & 0 & \cdots & 0\\ & p_{ij} & & \\ 0 & 0 & \cdots & 1\end{pmatrix}.$$

**例 3.4** 带反射壁的随机徘徊.

在上例中，若考虑在 $\xi$ 到 $a$ 或 $b$ 时反弹回一格，即

$$\zeta_n\triangleq\begin{cases}\xi_n(\omega), & \xi_{n-1}(\omega)\neq a,b;\\ a+1, & \xi_{n-1}=a;\\ b-1, & \xi_{n-1}=b.\end{cases}$$

不难证明 $\zeta\triangleq\{\zeta_n;n\geqslant0\}$ 也是一个马氏链，它的转移阵是

$$\boldsymbol{P}=\begin{pmatrix}0 & 1 & 0 & \cdots & 0 & 0\\ & & p_{ij} & & & \\ 0 & 0 & \cdots & 0 & 1 & 0\end{pmatrix}.$$

## § 3.2 常返与非常返

关于马氏链的一个自然的问题是：当 $n$ 越变越大，$\xi_n(\omega)$ 的分布变化是否会有什么倾向呢？首先我们以简单随机徘徊这种简单情形为例来考查这一问题.

### 1. 简单随机徘徊的常返与非常返

由强大数律，我们可以得到：当 $n\to+\infty$ 时，

$$\frac{1}{n}\xi_n = \frac{\xi_0+\sum_{k=1}^{n}x_k}{n} \to E(x_k) = p-q \quad (\text{a.e.}).$$

可见，对 $p>q$ 时，

$$P(\omega;\ \lim_{n\to\infty}\xi_n(\omega) = +\infty) = 1.$$

同理对 $p<q$ 时

$$P(\omega;\ \lim_{n\to\infty}\xi_n(\omega) = -\infty) = 1.$$

从而当 $p\neq q$ 时，对任意的 $|y|<+\infty$，

$$\begin{aligned}
0 &\leqslant P_y(\omega;\ \xi_n(\omega) = y,\ \text{i.o.}) \\
&\leqslant P_y(\omega;\ \text{存在 } n_k\to+\infty\ (\text{当 } k\to+\infty)\ \text{使 } \xi_{n_k} = y) \\
&\leqslant 1 - P(\omega;\ \lim_{n\to\infty}\xi_n(\omega) = +\infty \text{ 或 } -\infty) = 0,
\end{aligned}$$

其中 i.o. 表示其左边事件对无穷多个 $n$ 发生，$P_y$ 表示 $P(\cdot\mid\xi_0=y)$. 所以

$$P_y(\omega;\xi_n(\omega) = y,\quad \text{i.o.}) = 0. \tag{3.9}$$

现在考虑 $p=q$ 的情形. 由 § 2.1.4 段结果：

$$P_x(\omega;\ T_a(\omega) < +\infty) = 1\ (\text{对任意的 } a \text{ 与 } x).$$

于是对任意的 $y$，有限时间必离开 $y$ 达另一状态；此后又必再达 $y$，…如此无限地重复，可见

$$P_y(\omega;\xi_n(\omega) = y,\quad \text{i.o.}) = 1. \tag{3.10}$$

## 2. 马氏链的常返与非常返

上段的(3.9)与(3.10)分别表示状态 $y$ 的暂态性与常返性. 下面,我们给出一般定义:

**定义 3.4** 对马氏链 $\xi=\{\xi_n(\omega):n\geq 0\}$,状态 $y$ 称为**常返的**,如果(3.10)式成立. 若任意的状态 $y$ 都常返,则 $\xi$ 称为**常返的**;状态 $y$ 称为**暂态的**,如果(3.9)式成立. 如 $\xi$ 的任意状态均为暂态的,则称 $\xi$ 为**暂态的**. 暂态也可称非常返.

$R^1$ 上简单随机徘徊是常返的,当且仅当 $p=q=\frac{1}{2}$.

事实上马氏链的一个状态如不常返,就一定是暂态的,这是因为我们有下面的 0-1 律.

**命题 3.3** 设 $\xi$ 是一个马氏链,若记

$$\alpha=P_y(\omega:\xi_n(\omega)=y,\quad \text{i.o.}),$$

则 $\alpha$ 只能等于 0 或 1,即 $\alpha=\alpha^2$. 即 $\xi$ 不是常返的就必须是暂态的.

**证明** 令 $t_m(\omega)$ 为 $\xi$ 第 $m$ 次到达 $y$ 的时刻. 于是由马氏性

$$\begin{aligned}
\alpha &= P_y(\omega:\ \xi_n(\omega)=y,\quad \text{i.o.}) \\
&= \sum_{k=m}^{\infty} P_y(\omega:\ t_m(\omega)=k, \xi_k(\omega)=y; \xi_{k+n}(\omega)=y,\ \text{i.o.}) \\
&= \sum_{k=m}^{\infty} P_y(\omega:\ t_m(\omega)=k) P_y(\omega: \xi_n(\omega)=y,\ \text{i.o.}) \\
&\leq \sum_{k=m}^{\infty} P_y(\omega:\ \xi \text{ 必须至少到 } y\ \ m \text{ 次且 } t_m(\omega)=k)\cdot\alpha \\
&\leq P_y(\omega:\ \xi \text{ 必须至少到 } y\ \ m \text{ 次})\cdot\alpha \quad (m\to\infty) \\
&\to P_y(\omega:\ \xi \text{ 必须到 } y \text{ 无穷多次})\cdot\alpha=\alpha^2.
\end{aligned}$$

于是 $\alpha^2\leq\alpha\leq\alpha^2$,即 $\alpha=\alpha^2$. 从而 $\alpha(1-\alpha)=0$,即 $\alpha=0$ 或 1. 从而可见若(3.9)不成立,则(3.10)必成立. 也就是说不常返就必须是暂态的,从而我们也将暂态称为非常返.

**命题 3.4** 若 $P_x(\omega: T_y(\omega) < +\infty) = 1$，则当 $y$ 常返时

$$P_y(\omega: \xi_n(\omega) = y, \quad \text{i.o.}) = 1.$$

而当 $y$ 非常返时，则

$$P_x(\omega: \xi_n(\omega) = y, \quad \text{i.o.}) = 0.$$

**证明** 当 $y$ 非常返时，利用马氏性

$$\begin{aligned}
&P_x(\omega: \xi_n(\omega) = y, \quad \text{i.o.}) \\
&= \sum_k P_x(\omega: T_y(\omega) = k, \xi_k(\omega) = y, \xi_{k+n}(\omega) = y, \text{ i.o.}) \\
&= \sum_k P_x(\omega: T_y(\omega) = k) P_y(\omega: \xi_n(\omega) = y, \text{ i.o.}) \\
&= P_x(\omega: T_y(\omega) < +\infty) P_y(\omega: \xi_n(\omega) = y, \text{ i.o.}) \\
&= 0 \quad (\text{由命题 } 3.3).
\end{aligned}$$

同理当常返 $y$ 时

$$\begin{aligned}
&P_x(\omega: \xi_n(\omega) = y, \text{ i.o.}) \\
&\quad = P_x(\omega: T_y(\omega) < +\infty) P_y(\omega: \xi_n(\omega) = y, \text{ i.o.}) \\
&\quad = 1.
\end{aligned}$$

### 3. 马氏链的状态分类

在 § 3.1 例 3.3 中的边界状态 $a$ 与 $b$ 和其他状态 $(\{i; a < i < b\})$ 是很不同的. 由于

$$P_x(\omega: \text{在有限时间 } \xi(\omega) \text{ 必达 } a \text{ 或 } b) = 1,$$

$$P_a(\omega: \xi_n(\omega) = a, \text{对 } \forall\, n \geqslant 1) = 1,$$

$$P_b(\omega: \xi_n(\omega) = b, \text{对 } \forall\, n \geqslant 1) = 1,$$

我们可以看出边界 $a, b$ 都是常返状态，但是对 $x \neq a, b$. 我们有

$$\begin{aligned}
&P_x(\omega: \xi_n(\omega) = x, \text{ i.o.}) \\
&\leqslant 1 - P_x(\omega: \xi_n(\omega) \text{ 在有限时间达 } a \text{ 或 } b, \text{并永停留在那里}) \\
&= 1 - 1 = 0.
\end{aligned}$$

可见这里的非边界状态 $x$ 是非常返的.

又例如在例 3.4 中如考虑简单情形 $a = 1, b = 3$

$$P=\begin{pmatrix}0&1&0\\q&0&p\\0&1&0\end{pmatrix},$$

则容易由归纳法验证

$$P^{2n+1}=\begin{pmatrix}0&1&0\\q&0&p\\0&1&0\end{pmatrix},\qquad P^{2n}=(p_{ij}(2n))=\begin{pmatrix}q&0&p\\0&1&0\\q&0&p\end{pmatrix}.$$

在这种情况下，无论是否是边界状态，都是常返的. 下面我们来说明这一点.

对 $x=2$，由于 $p_{22}(2n)\equiv1(n\geqslant1)$，可见

$$P_2(\omega:\xi_{2n}(\omega)=2,\text{对}\ \forall\ n\geqslant1)=1,$$

即
$$P_2(\omega:\xi_n(\omega)=2,\text{i.o.})=1.$$

对 $x=1,3$（边界点），我们有

$$\begin{aligned}&P_1(\omega:T_1(\omega)=2k)\\&\quad=P_1(\omega:\text{对}\ \forall\ 1\leqslant m<k,\xi_{2m}=3,\xi_{2m+1}=2,\xi_{2k}=1)\\&\quad=1\cdot p\cdot1\cdot p\cdots1\cdot p\cdot1\cdot q\\&\quad=p^{k-1}\cdot1^k\cdot q=p^{k-1}\cdot q.\end{aligned}$$

同理

$$P_3(\omega:T_3(\omega)=2k)=q^{k-1}\cdot p.$$

于是

$$P_1(\omega:T_1(\omega)<+\infty)=\sum_{k=1}^{\infty}q^{k-1}\cdot p=p\cdot\frac{1}{1-q}=1,$$

$$P_3(\omega:T_3(\omega)<+\infty)=\sum_{k=1}^{\infty}p^{k-1}\cdot q=1.$$

也就是说从边界状态“1”与“3”出发总是要在有限时间回来的，因而以概率为1地不断重复这个回来，离开；再回来，再离开的过程，$\xi$ 必定无穷次回到它出发的状态.

从上面的例子可以看出，状态是否常返并不一定要求过程以

概率为 1 地留在原地，而与所有可以到达的状态的性质有关. 为此在本节中我们将对状态按它们是否相通来分类，并进而考查这样分类的常返性.

**定义 3.5（可达与互通）** 状态 $i$ 称为可达状态 $j$（记为 $i\to j$），如果存在 $i=i_0,i_1,\cdots,i_n,i_{n+1}=j$ 使得

$$p_{i_k i_{k+1}}>0 \quad (\text{对任意的 } k=0,1,\cdots,n \text{ 成立}).$$

状态 $i$ 与 $j$ 称为**互通**，记为 $i\sim j$，如果 $i\to j$ 且 $j\to i$. 此外，我们规定 $i\sim i$.

这样，互通关系是一个等价关系，即满足：

1）$i\sim j$ 蕴含 $j\sim i$；

2）$i\sim j$，$j\sim k$ 蕴含 $i\sim k$；

3）$i\sim i$.

于是，可按关系～将状态空间 $\mathscr{S}$ 分类：$\mathscr{S}=\bigcup\limits_n E_n$，每类的全部状态互通，而且在此类之外，再无其他状态能与类中的状态互通. 例如在例 3.3 中

$$\mathscr{S}=\{a,a+1,\cdots,b-1,b\}=\{a\}\cup\{b\}\cup\{i;a<i<b\},$$

其中 $\{a\}$，$\{b\}$ 两类中分别只含一个点（边界点）. 在例 3.4 中全部状态都互通，因而只有一个互通类.

在下文中我们将看到，这种互通类的分法，对于常返态至关重要. 而对于暂态则并无重要作用，因此许多书中对于暂态根本不予分类，而让它们自成一类——暂态类.

状态空间 $\mathscr{S}$ 的子集 $A$ 称为 ***P*-闭集**，如果

$$\sum_{j\in A}p_{ij}=1 \quad (\forall\, i\in A);$$

如果 $\mathscr{S}$ 无真 $P$-闭子集，则称马氏链的转移阵 $P$ 为不可约的，这时我们也简称马氏链 $\xi$ 为**不可约的**.

**命题 3.5** 若 $i\to j$，$i$ 常返，则

$$P_j(\omega;T_i(\omega)<+\infty)=1.$$

**证明** "$i$ 常返"蕴含"$P_j(\omega:\tilde{T}_i(\omega)<+\infty)=1$". 其中

$$T_i(\omega)\triangleq\inf\{n\geqslant 1;\xi_n(\omega)=i\},$$

$$\tilde{T}_i(\omega)=\begin{cases}T_i(\omega), & \text{当 } \xi_0(\omega)\neq i;\\ 0, & \text{当 } \xi_0(\omega)=i.\end{cases}$$

若令 $m_0=\min\{m:p_{ij}(m)>0\}$, 则

$$p_{ij}(m)=0\quad(m<m_0),\qquad p_{ij}(m_0)>0.$$

因而

$$\begin{aligned}&P_i(\omega:\xi_{m_0}(\omega)=j,\xi_m(\omega)=i,1\leqslant m\leqslant m_0)\\ &\leqslant\sum_{1\leqslant m<m_0}p_{ii}(m)p_{ij}(m_0-m)=0.\end{aligned}$$

于是

$$\begin{aligned}1&=P_i(\omega:\tilde{T}_i(\omega)<+\infty)\\ &=P_i(\omega:\xi_{m_0}(\omega)=j,\tilde{T}_i(\omega)<+\infty)\\ &\quad+P_i(\omega:\xi_{m_0}(\omega)\neq j,\tilde{T}_i(\omega)<+\infty)\\ &\leqslant P_i(\omega:\xi_{m_0}(\omega)=j,m_0<\tilde{T}_i(\omega)<+\infty)+\sum_{k\neq j}p_{ik}(m_0)\\ &=P_i(\omega:\xi_{m_0}(\omega)=j,m_0\text{ 之后 }\xi\text{ 必达 }i)+(1-p_{ij}(m_0))\\ &=p_{ij}(m_0)P_j(\omega:T_i(\omega)<+\infty)+1-p_{ij}(m_0)\\ &=1-p_{ij}(m_0)(1-P_j(\omega:T_i(\omega)<+\infty))\\ &\leqslant 1.\end{aligned}$$

可见 $1-P_j(\omega:T_i(\omega)<+\infty)=0$, 即由 $j$ 出发有限时间必达 $i$.

**命题 3.6** 状态 $i$ 常返等价于以下两条件之一:

(1) $P_i(\omega:\tilde{T}_i(\omega)<+\infty)=1$;

(2) $\sum_n p_{ii}(n)=+\infty$.

**证明** "$i$ 常返"等价于"从 $i$ 出发以概率为 1 地有限时间内必

返回 $i$”. 这是因为时齐马氏链概率为 1 地返回 1 次,就有 2 次,3 次,…,无穷多次. 从而“$i$ 常返”等价于(1).

另一方面,令

$$f_{ij}(n) \triangleq P_i(\omega: T_j(\omega) = n) = P_i(\omega: \xi \text{ 在 } n \text{ 首达 } j). \tag{3.11}$$

于是

$$P_i(\omega: T_j(\omega) < +\infty) = \sum_{n=1}^{\infty} f_{ij}(n) \triangleq f_{ij}^*. \tag{3.11$'$}$$

此外,对 $j \neq i$

$$\begin{aligned}
p_{ij}(n) &= P_i(\omega: \xi_n(\omega) = j) \\
&= \sum_{m=1}^{n} P_i(\omega: T_j(\omega) = m; \xi_n(\omega) = j) \\
&= \sum_{m=1}^{n} P_i(\omega: T_j(\omega) = m) P(\xi_n(\omega) = j \mid \xi_m = j) \\
&= \sum_{m=1}^{n} f_{ij}(m) p_{jj}(n-m).
\end{aligned} \tag{3.12}$$

对固定的 $i$,看幂级数

$$F(z) = \sum_{m=1}^{\infty} f_{ii}(m) z^m \ (|z| \leqslant 1), \tag{3.13}$$

$$P(z) = \sum_{n=0}^{\infty} p_{ii}(n) z^n \ (|z| < 1). \tag{3.14}$$

于是

$$\begin{aligned}
P(z) &= \sum_{n=1}^{\infty} \sum_{m=1}^{n} f_{ii}(m) p_{ii}(n-m) z^n + 1 \\
&= \sum_{m=1}^{\infty} \left( \sum_{n=m}^{\infty} p_{ii}(n-m) z^{n-m} \right) f_{ii}(m) z^m + 1 \\
&= \sum_{m=1}^{\infty} f_{ii}(m) z^m \sum_{k=0}^{\infty} p_{ii}(k) z^k + 1 \\
&= F(z) P(z) + 1.
\end{aligned}$$

从而

$$P(z)=\frac{1}{1-F(z)}\quad(|z|<1).$$

令 $z\uparrow 1$，可见 $\sum_n p_{ii}(n)$ 收敛当且仅当 $\sum_n f_{ii}(n)=1$. 此即(1)与(2)等价. 命题证毕.

**推论** 状态 $i$ 非常返当且仅当 $\sum_n p_{ii}(n)<+\infty$.

由于命题 3.3，推论显然.

**命题 3.7** 若 $i$ 常返，$i\to j$，则 $j\to i$，而且 $j$ 也常返.

**证明** 命题 3.5 说明 $j\to i$. 于是 $\exists\, n_0, m_0$ 使

$$p_{ji}(n_0)>0,\ p_{ij}(m_0)>0,$$

从而 $$p_{jj}(n+n_0+m_0)\geqslant p_{ji}(n_0)p_{ii}(n)p_{ij}(m_0).$$

于是由 $\sum_n p_{ii}(n)$ 发散可见 $\sum_n p_{jj}(n)$ 也发散. 再由命题 3.6 知 $j$ 也是常返状态.

从命题 3.7 可见一个前面所讲的互通类中有一个状态常返，则其中任一个状态都常返. 这意味着常返是整个互通类的性质. 再则，不可约链或者是只有一个互通类的常返链，或者全是暂态.

## §3.3 马氏链的转移概率的极限与不变分布

本节中我们给出关于马氏链的最重要的结果之一——关于 $n$ 步转移概率矩阵 $\boldsymbol{P}^n$，当 $n\to+\infty$ 时的极限的研究，以及极限与不变分布的关系.

### 1. $n$ 步转移概率矩阵 $\boldsymbol{P}^n$ 的极限研究

首先我们对一个特殊情况——全部状态可在一步互通的有限状态马氏链，给出 $\boldsymbol{P}^n$ 的极限.

**定理 3.8** 设 $\xi$ 是一个马氏链，它的状态空间 $\mathscr{S}$ 只有有限个元素(不妨设 $\mathscr{S}=\{1,2,\cdots,N\}$)，而且对 $\mathscr{S}$ 中任意二个状态 $i$ 与

$j$，都有 $p_{ij}>0$，则存在 $\mathscr{S}$ 上的一个分布 $\pi=\{\pi_1,\pi_2,\cdots,\pi_N\}$ 使得对一切 $\mathscr{S}$ 中的状态 $i$ 与 $j$ 都有：

1) $\sum\limits_k \pi_k p_{kj}=\pi_j, \pi_j\geqslant\delta$；

2) $|p_{ij}(n)-\pi_j|\leqslant(1-N\delta)^n$，其中 $\delta\triangleq\min\{p_{ij}: i,j\in\mathscr{S}\}\leqslant\frac{1}{N}$.

注：满足 2)的 $P$ 称为具有指数遍历性.

**证明** 令 $M_j^{(n)}$ 与 $m_j^{(n)}$ 分别是矩阵 $\boldsymbol{P}^n$ 的第 $j$ 列的最大与最小值，显然

$$\delta\triangleq\min\{p_{ij}: i,j\in\mathscr{S}\}\leqslant p_{ij}\leqslant 1-\sum_{k\neq j}p_{ik}\leqslant 1-(N-1)\delta,$$

因而

$$\delta\leqslant\frac{1}{N}\sum_{k=1}^{N}p_{ik}=\frac{1}{N},\quad m_j^{(1)}\geqslant\delta,\quad M_j^{(1)}\leqslant 1-(N-1)\delta.$$

从而

$$M_j^{(1)}-m_j^{(1)}\leqslant 1-(N-1)\delta-\delta=1-N\delta.\quad(3.15)$$

而且

$$\begin{aligned}M_j^{(n+1)}&=\max_i p_{ij}(n+1)=\max_i\sum_k p_{ik}p_{kj}(n)\\&\leqslant\max_i\sum_k p_{ik}M_j^{(n)}=M_j^{(n)}.\end{aligned}$$

类似地，$m_j^{(n+1)}=\min\limits_i p_{ij}(n+1)\geqslant\min\limits_i\sum\limits_k p_{ik}m_j^{(n)}=m_j^{(n)}$，也就是说 $M_j^{(n)}$ 是单调下降的，而 $m_j^{(n)}$ 是单调上升的，如果我们还能进而说明

$$0\leqslant M_j^{(n)}-m_j^{(n)}\leqslant(1-N\delta)^n\to 0\quad(n\to+\infty),(3.16)$$

就能推出各列当 $n\to+\infty$ 时趋于相同极限. 进而若令 $\pi_j\triangleq\lim\limits_{n\to\infty}M_j^{(n)}$，则由

$$m_j^{(n)}-M_j^{(n)}\leqslant p_{ij}(n)-\pi_j\leqslant M_j^{(n)}-m_j^{(n)}$$

就知道 $\lim\limits_{n\to\infty}p_{ij}(n)=\pi_j$，从而有

$$\pi_j = \lim_{n\to\infty} p_{ij}(n+1) = \lim_{n\to\infty}\sum_{k=1}^{N} p_{ik}(n)p_{kj} = \sum_{k=1}^{N}\pi_k p_{kj}.$$

可见，为证明定理只要证明(3.16)即可．为此注意

$$\begin{aligned}
M_j^{(n+1)} - m_j^{(n+1)} &= \max_i p_{ij}(n+1) - \min_i p_{ij}(n+1) \\
&= \max_i \sum_k p_{ik}p_{kj}(n) - \min_i \sum_k p_{ik}p_{kj}(n) \\
&\triangleq \sum_k p_{i_1k}p_{kj}(n) - \sum_k p_{i_2k}p_{kj}(n) \\
&= \sum_k{}^{+} (p_{i_1k} - p_{i_2k})p_{kj}(n) - \sum_k{}^{-} (p_{i_2k} - p_{i_1k})p_{kj}(n) \\
&\leqslant M_j^{(n)}\sum_k{}^{+} (p_{i_1k} - p_{i_2k}) - m_j^{(n)}\sum_k{}^{-} (p_{i_2k} - p_{i_1k}),
\end{aligned} \tag{3.17}$$

又由于

$$\begin{aligned}
0 &= \sum_k{}^{+} (p_{i_1k} - p_{i_2k}) + \sum_k{}^{-} (p_{i_1k} - p_{i_2k}) \\
&= \sum_k{}^{+} (p_{i_1k} - p_{i_2k}) - \sum_k{}^{-} (p_{i_2k} - p_{i_1k}).
\end{aligned}$$

于是

$$\begin{aligned}
\sum_k{}^{-} (p_{i_2k} - p_{i_1k}) &= \sum_k{}^{+} (p_{i_1k} - p_{i_2k}) \\
&= \Big(1 - \sum_k{}^{-} p_{i_1k}\Big) - \sum_k{}^{+} p_{i_2k} \\
&\leqslant 1 - \sum_k{}^{-} \delta - \sum_k{}^{+} \delta = 1 - N\delta.
\end{aligned} \tag{3.18}$$

从而(3.17)式就变为

$$M_j^{(n+1)} - m_j^{(n+1)} \leqslant (M_j^{(n)} - m_j^{(n)})\sum_k{}^{+} (p_{i_1k} - p_{i_2k}).$$

这样由归纳法，从(3.18)与(3.15)就可得到(3.16)．定理证毕．

定理 3.8 及其证明说明了随着 $n$ 的增长，$\boldsymbol{P}^n$ 的每一列（例如第 $j$ 列）各元素越来越接近，最终以行向量 $\underline{\pi}$ 的第 $j$ 个分量 $\pi_j$ 为共同极限．于是 $\boldsymbol{P}^n$ 的各行当 $n\to+\infty$ 时，均以 $\underline{\pi}$ 为极限．这意味着无论从什么初分布出发，当时间充分长后 $\xi_n$ 的分布渐近于 $\underline{\pi}$：

$$\lim_{n\to\infty} P_i(\omega:\xi_n(\omega)=j)=\lim_{n\to\infty} p_{ij}(n)=\pi_j,$$

而且

$$\lim_{n\to\infty} P(\omega:\xi_n(\omega)=j)=\lim_{n\to\infty}\sum_{i=1}^{N} P(\omega:\xi_0(\omega)=i)P_i(\omega:\xi_n(\omega)=j)$$

$$=\sum_{i=1}^{N} P(\omega:\xi_0(\omega)=i)\lim_{n\to\infty} p_{ij}(n)$$

$$=\sum_{i=1}^{N} P(\omega:\xi_0(\omega)=i)\pi_j=\pi_j.$$

**2. 不变分布与 $P^n$ 的极限(如存在)的求法**

在第一段中，$P^n$ 每一行的极限 $\underline{\pi}$ 是一个 $\mathscr{S}$ 上的分布，这是因为：

$$\pi_j=\lim_{n\to\infty} p_{ij}(n)\geqslant 0,$$

$$\sum_{j=1}^{N}\pi_j=\sum_{j=1}^{N}\lim_{n\to\infty} p_{ij}(n)=\lim_{n\to\infty}\sum_{j=1}^{N} p_{ij}(n)=1.$$

又由于它满足

$$\sum_k \pi_k p_{kj}=\pi_j\quad(\forall\, j\in\mathscr{S}),\tag{3.19}$$

容易从归纳法及

$$\sum_k \pi_k p_{kj}(n+1)=\sum_k \pi_k\sum_i p_{ki}p_{ij}(n)$$

$$=\sum_i\Big(\sum_k \pi_k p_{ki}\Big)p_{ij}(n)=\sum_i \pi_i p_{ij}(n)$$

得到

$$\sum_k \pi_k p_{kj}(n)=\pi_j.\tag{3.20}$$

于是我们得到了以下命题：

**命题 3.9** 若 $\underline{\pi}$ 满足(3.19)，则必对一切 $n\geqslant 1$，满足(3.20)。从而，若 $\xi$ 的初分布为 $\pi$ (即 $P(\omega:\xi_0(\omega)=i)=\pi_i$ )就有

$$P(\omega:\xi_n(\omega)=j)=\sum_i \pi_i p_{ij}(n)=\pi_j.$$

可见从初分布 $\underline{\pi}$ 出发，$\xi$ 任意时刻的分布都仍为 $\underline{\pi}$.

我们给出以下定义：

**定义 3.6(不变分布)** 对以 $\boldsymbol{P}$ 为转移矩阵，$\mathscr{S}$ 为状态空间的时齐马氏链 $\xi$，概率分布 $\pi=\{\pi_i;\quad i\in\mathscr{S}\}$ 称其为**不变概率分布**，如果

$$\pi\boldsymbol{P}=\pi\quad\left(\text{即}\sum_k\pi_k p_{kj}=\pi_j\text{ 对 } j\in\mathscr{S}\text{ 成立}\right).\qquad(3.21)$$

**命题 3.10** 若定义 3.6 中的时齐马氏链 $\xi$ 以其不变分布为初分布，则对任意的 $k\geqslant1,s\geqslant1$，及时刻 $n_1,\cdots,n_k$，$\{\xi_{n_1}(\omega),\cdots,\xi_{n_k}(\omega)\}$ 与 $\{\xi_{n_1+s}(\omega),\cdots,\xi_{n_k+s}(\omega)\}$ 永远同分布，从而 $\xi$ 与 $\tilde{\xi}$ 也同分布，其中 $\tilde{\xi}\triangleq\{\xi_{n+s};n=0,1,\cdots\}$，也就是 $\xi$ 的分布对时间平移不变.

**证明**

$$\begin{aligned}
&P(\omega:\xi_{n_1}(\omega)=i_1,\cdots,\xi_{n_k}(\omega)=i_k)\quad(\text{由命题 3.2})\\
&\quad=P(\omega:\xi_{n_1}(\omega)=i_1)p_{i_1i_2}(n_1-n_2)\cdots p_{i_{k-1}i_k}(n_k-n_{k-1})\\
&\quad=\pi_{i_1}p_{i_1i_2}(n_1-n_2)\cdots p_{i_{k-1}i_k}(n_k-n_{k-1})\\
&\quad=P(\omega:\xi_{n_1+s}(\omega)=i_1,\cdots,\xi_{n_k+s}(\omega)=i_k)\quad(\text{由命题 3.2}).
\end{aligned}$$

**命题 3.11** 若 $\lim\limits_{n\to\infty}p_{ij}(n)=\pi_j$ 对 $\forall\, i,j\in\mathscr{S}$ 成立，且 $\sum\limits_j\pi_j=1$，则 $\underline{\pi}$ 是 $\boldsymbol{P}$（或说 $\xi$）的唯一不变概率分布. 即 $\pi$ 是方程

$$\begin{cases}\underline{\pi}P=\underline{\pi},\\ \sum\limits_i\pi_i=1\end{cases}\qquad(3.22)$$

的非负解. 又若 $\xi(P)$ 不可约，则(3.22)的非负解存在唯一. $\pi$ 也是矩阵 $\boldsymbol{P}$ 的特征值为 1 的左特征向量，而且 $\boldsymbol{P}$ 的特征值为 1 的特征空间是一维的.

**证明** 这里我们仅对定理 3.8 条件成立的情形证明. 这样我们有

$$\pi_j=\lim_{n\to\infty}p_{ij}(n+1)=\lim_{n\to\infty}\sum_k p_{ik}(n)p_{kj}=\sum_k \mu_k p_{kj}.$$

设 $x=(x_1,\cdots,x_N)$ 是 $xP=x,\sum_i x_i=1$ 的解，则由定理 3.8

$$x_j=\sum_k x_k p_{kj}(n)\xrightarrow[n\to\infty]{}\sum_k x_k\pi_j=\pi_j\sum_k x_k=\pi_j.$$

命题证毕.

命题 3.11 告诉我们只要解方程(3.22)就能求出 $\lim\limits_{n\to\infty}p_{ij}(n)$ (如果极限存在).

**例 3.5** 设 $P^2=\begin{pmatrix}\frac{1}{6}&\frac{1}{2}&\frac{1}{3}\\ \frac{1}{3}&\frac{1}{6}&\frac{1}{2}\\ \frac{1}{2}&\frac{1}{3}&\frac{1}{6}\end{pmatrix}$，则 $\pi$ 是下面方程的唯一解：

$$\begin{cases}\frac{5}{6}x_1-\frac{1}{2}x_2-\frac{1}{3}x_3=0,\\ -\frac{1}{3}x_1+\frac{5}{6}x_2-\frac{1}{2}x_3=0,\\ -\frac{1}{2}x_1-\frac{1}{3}x_2+\frac{5}{6}x_3=0,\\ x_1+x_2+x_3=1.\end{cases}$$

容易得到上面方程的唯一解是

$$(x_1,x_2,x_3)=\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right),$$

从而由定理 3.8 知道以 $\boldsymbol{P}$ 为转移阵的时齐马氏链 $\xi$ 应满足：

$$\lim_{n\to\infty}P_i(\xi_n(\omega)=j)=\frac{1}{3}\quad(\forall\ i,j=1,2,3).$$

也就是说时间充分长以后，$\xi_n(\omega)$ 近于均匀分布.

**例 3.6** 设有 6 个车站，它们之间有公路连接，情况如图. 如果汽车每天可以从一个站驶到与之直接有公路相连结的相邻车站，在夜间到达车站留宿并接受加油、清洗、检修等服务，次日清晨

各站按相同的比例将各汽车派往其相邻车站．试说明在运行了很多日子以后，各站每晚留宿的汽车比例趋于稳定，并求出这个比例，以便正确地设置各站的服务规模．

**解**　设一个汽车在第 $n$ 天留宿的站号为 $\xi_n(\omega)$，于是在已知 $\xi_n(\omega)=k$ 的条件下，第 $n+1$ 天它留宿的站号是以 $\frac{1}{n_k}$ 的概率取与 $k$ 站相邻的各站之一（$n_k$ 为与 $k$ 站相邻的车站数），而与第 $n-1$ 天及其以前的各天汽车留宿何站无关．从而一个汽车的留宿（运行）情况：$\{\xi_n(\omega):n=0,1,\cdots\}$ 就是一个时齐马氏链，其转移矩阵是：

$$P=\begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \end{pmatrix}.$$

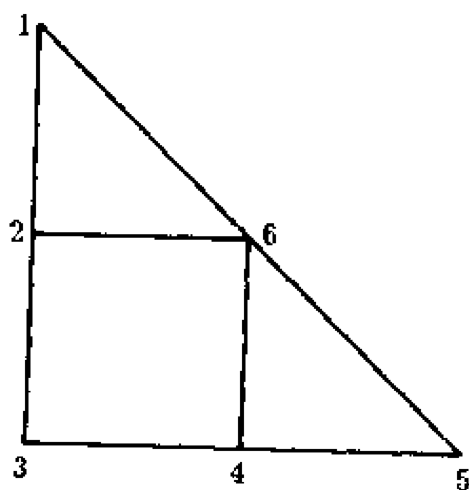


图　3.1

解方程

$$\begin{cases} \underline{\pi} P = \underline{\pi}, \\ \sum_{i=1}^{6} \pi_i = 1, \end{cases}$$

可得到 $\underline{\pi} = \left\{ \frac{1}{8}, \frac{3}{16}, \frac{1}{8}, \frac{3}{16}, \frac{1}{8}, \frac{1}{4} \right\}$.

值得指出的是照常规直接求解方程 $\underline{\pi} P = \underline{\pi}$ 是比较麻烦的，下面我们给出的定理 3.21 及可逆马氏链理论，可以大大简化这种计算(请参见下面 § 3.5 例 3.9).

### 3. $P^n$ 的平均极限

在本节前两段中，指出了一步互通的时齐马氏链的转移矩阵 $P^n$ 有一个各行均相同的极限. 但是一般地讲 $P^n$ 并不一定有极限. 例如当

$$P = \begin{pmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{pmatrix},$$

则

$$P^2 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \qquad P^{2n+1} = P.$$

可见 $p_{11}(2n)=\frac{1}{2}, p_{11}(2n+1)=0, \lim\limits_{n} p_{11}(n)$ 不存在. 这时定理 3.9 一般不再成立,但我们有下面的定理:

**定理 3.12** 设 $\boldsymbol{P}$ 是一个 $N \times N$ 转移矩阵,则

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}(k) (\triangleq L_{ij})$$

一定存在. 又令 $L=(L_{ij})$,则以下矩阵运算等式成立:

$$\boldsymbol{L}\boldsymbol{P}^m = \boldsymbol{P}^m \boldsymbol{L} = \boldsymbol{L} = \boldsymbol{L}^2 \quad (\forall\ m \geqslant 1).$$

**证明** 为避免技术性的困难,我们这里只对有限状态($N$ 个状态)马氏链来证明. 但本定理对可数状态也成立. 令

$$L_{ij}(n) \triangleq \frac{1}{n} \sum_{k=1}^{n} p_{ij}(k), \quad 0 \leqslant L_{ij}(n) \leqslant 1.$$

对 $N^2$ 个序列 $\{L_{ij}(n); n=0,1,\cdots\}$ 一定可以找到公共的 $\{n_k, k=1, 2,\cdots\}; n_k \to +\infty$,使它们各自的子列有:

$$L_{ij}(n_k) \to L_{ij} \quad (\text{当 } k \to +\infty).$$

又由于

$$\begin{aligned}
\sum_{l} L_{il}(n_k) p_{lj}(m) &= \sum_{l} \frac{1}{n_k} \sum_{r=1}^{n_k} p_{il}(r) p_{lj}(m) \\
&= \frac{1}{n_k} \sum_{r=1}^{n_k} \sum_{l} p_{il}(r) p_{lj}(m) = \frac{1}{n_k} \sum_{r=1}^{n_k} p_{ij}(m+r) \\
&= \frac{1}{n_k} \left\{ \sum_{r=1}^{n_k} p_{ij}(r) - \sum_{r=1}^{m} p_{ij}(r) + \sum_{r=1}^{m} p_{ij}(n_k + r) \right\},
\end{aligned}$$

其中第一项以 $L_{ij}$ 为极限,第二三两项都绝对值小于 $\frac{m}{n_k}$,因而极限均为 0,从而

$$\begin{aligned}
\sum_{l=1}^{N} L_{il} p_{lj}(m) &= \sum_{l=1}^{N} \lim_{k\to\infty} L_{il}(n_k) p_{lj}(m) \\
&= \lim_{k\to\infty} \left\{ \sum_{l=1}^{N} L_{il}(n_k) p_{lj}(m) \right\} = L_{ij}.
\end{aligned}$$

此即

$$\boldsymbol{L}\boldsymbol{P}^m = \boldsymbol{L} \quad (m \geqslant 1).$$

同理 $$P^m L = L \quad (m \geqslant 1).$$

于是 $$L\left(\frac{1}{n}\sum_{m=1}^{n} P^m\right) = L \quad (m \geqslant 1).$$

若记 $L(n) \triangleq (L_{ij}(n))$，则我们得到 $LL(n) = L$，即

$$\sum_{l=1}^{N} L_{il} L_{lj}(n) = L_{ij}. \tag{3.23}$$

由此我们断言 $\lim_{n\to\infty} L_{ij}(n)$ 存在. 事实上，如果 $L_{ij}(n)$ 另有一个收敛子列 $L_{ij}(m_k)$ 收敛到某个 $\tilde{L}_{ij}$. 那么由(3.23)可推出 $L \cdot \tilde{L} = L$ 及 $\tilde{L} \cdot L = L$. 同理又有 $\tilde{L} \cdot L = L$，从而 $\tilde{L} = L$，即极限存在，而且

$$L^2 = L \cdot \tilde{L} = L.$$

定理证毕.

在第一段中，$L$ 是每行一样的，那么一般地 $L$ 应该是什么样子的呢？若 $\xi$(或 $P$) 只有一个互通常返类，也即它的任意两个状态都互通，则 $L$ 一定是各行都相同的矩阵. 下面的定理 3.13 给出了这一结果.

**定理 3.13** 对任意状态 $i$ 与 $j$，我们都有

$$L_{ij} = \left(\sum_k f_{ij}(k)\right) L_{jj}. \tag{3.24}$$

又设 $\xi$(转移阵为 $P$)是一个互通常返时齐马氏链，则

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} p_{ij}(k) = L = \mathbf{1}\pi,$$

其中 $\pi$ 是 $P$ 的不变概率分布，$\mathbf{1}$ 是一个全部元素均为 1 的列向量. 在有限状态情形，互通常返时齐马氏链的不变概率测度是唯一的.

**证明** 我们仍然只对有限状态的情形证明定理.

1) 先证明(3.24)及 $\xi$ 不可约常返时，$L$ 的各行均相同. 事实上，由于(3.12)

$$p_{ij}(k) = \sum_{l=1}^{k} f_{ij}(l) p_{jj}(k-l),$$

$$L_{ij}(n)=\frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{k}f_{ij}(l)p_{jj}(k-l)$$

$$=\sum_{l=1}^{n}f_{ij}(l)\frac{1}{n}\sum_{k=l}^{n}p_{jj}(k-l).$$

利用下面的引理 3.14，就得到

$$L_{ij}=\lim_{n\to+\infty}L_{ij}(n)=\left(\sum_{l=1}^{\infty}f_{ij}(l)\right)L_{jj}.$$

当 $\xi$ 不可约，常返时，由于

$$\sum_{l=1}^{\infty}f_{ij}(l)=P_i(\omega:T_j(\omega)<+\infty)=1.$$

从而
$$L_{ij}=L_{jj},$$

即 $\boldsymbol{L}$ 每行均相同，将其记为 $\underline{\pi}=(\pi_1,\cdots,\pi_n,\cdots)$，即 $\pi_j=L_{jj}$.

2）现在来证明 $\underline{\pi}$ 是不变概率分布. 这里我们仍然只对有限状态马氏链证明，以避免技术上的麻烦. 可数状态情形定理仍成立.

$$\pi_j=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}p_{ij}(k)=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\sum_{l}p_{il}(k-1)p_{lj}$$

$$=\sum_{l=1}^{N}\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}p_{il}(k-1)p_{lj}=\sum_{l=1}^{N}\lim_{n\to\infty}\frac{1}{n}\left(\sum_{k=1}^{n-1}p_{il}(k)+\delta_{il}\right)p_{lj}$$

$$=\sum_{l=1}^{N}\lim_{n\to\infty}\frac{n-1}{n}L_{il}(n-1)p_{lj}=\sum_{l=1}^{N}L_{il}p_{lj}=\sum_{l=1}^{n}\pi_l p_{lj}.$$

3）唯一性证明类似于命题 3.11，留给读者自己去证明.

**引理 3.14** 设 $a_n\geqslant 0$， $b_n\leqslant c$， $k\geqslant 0$， $b_k\geqslant 0$，而且 $\sum_n a_n=A<+\infty$，$\lim_{n\to\infty}b_{n,k}=b_k$，则

$$\lim_{n\to\infty}\sum_{k=1}^{n}a_k b_{n,k}=\sum_{k=1}^{\infty}a_k b_k.$$

证明见附录.

对于非常返状态，我们有以下结果：

**定理 3.15** 设 $j$ 是 $\xi$ 的非常返状态，则

$$\lim_{n\to\infty} p_{ij}(n) = 0.$$

**证明** 由命题 3.6 的推论

$$\sum_n p_{jj}(n) < +\infty.$$

从而
$$\lim_{n\to\infty} p_{jj}(n) = 0.$$

再由引理 3.14

$$\lim_{n\to\infty} p_{ij}(n) = \lim_{n\to\infty} \sum_{k=1}^{n} f_{ij}(k) p_{jj}(n-k) = 0.$$

由定理 3.15 与 3.13，我们就可以知道：

1) $\boldsymbol{L}$ 中对应于非常返状态 $j$ 的列的元素 $L_{ij} = 0$；

2) 当 $j$ 是常返状态，$i$ 是非常返状态

$$L_{ij} = \left( \sum_k f_{ij}(k) \right) L_{jj};$$

3) 对 $i$ 是常返状态，由命题 3.7，可见 $i$ 只可达与 $i$ 属于同一个不可约互通类的状态，即对任意 $n$ 有

$$p_{ij}(n) = 0 \text{ (当 } j \text{ 不在 } i \text{ 的常返互通类中)}.$$

综上所述，若 $\xi$ 的状态分为常返互通类 $H_1, \cdots, H_m, \cdots$ 及 $T$（全部非常返状态集），则

$$\boldsymbol{L} = \begin{array}{c} \\ T \\ H_1 \\ H_2 \\ \vdots \\ H_m \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} T & H_1 & H_2 & \cdots & H_m & \cdots \end{array} \\ \left[ \begin{array}{cccccc} 0 & \sum_k f_{ij}(k)\pi_j^{(1)} & \cdots & \cdots & \sum_k f_{ij}(k)\pi_j^{(m)} & \cdots \\ 0 & \underline{\mathbf{1}}\pi^{(1)} & 0 & \cdots & 0 & \cdots \\ 0 & 0 & \underline{\mathbf{1}}\pi^{(2)} & \cdots & 0 & \cdots \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots \\ 0 & 0 & 0 & \cdots & \underline{\mathbf{1}}\pi^{(m)} & \cdots \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots \end{array} \right] \end{array}.$$

由此可见研究 $\boldsymbol{P}^n$ 的平均极限只须限制在每个正常返互通类中研究.

下面我们给出转移概率平均极限的概率意义. 令

$$N_{j,n}(\omega) = \{m \leqslant n; \xi_m(\omega) = j\},$$

则

$$v_{n,j}(\omega) \triangleq \frac{N_{j,n}(\omega)}{n} = \frac{1}{n}\sum_{k=1}^{n} 1_{\{j\}}(\xi_k(\omega))$$

就表示在前 $n$ 次 $\xi$ 处在状态 $j$ 的频率，于是

$$L_{ij}(n) = E_i v_{n,j}(\omega), \quad L_{ij} = \lim_{n\to\infty} E_i v_{n,j}(\omega).$$

可见 $L_{ij}$ 可看作从 $i$ 出发，$\xi$ 逗留于 $j$ 的平均频率. 对于不可约常返马氏链，这个平均频率不依赖于出发状态 $i$，因而也不依赖于初分布.

#### 4. 关于周期与 $P^n$ 的极限的附注

在第三段一开始我们给出了 $P^n$ 极限不存在的反例，再对比定理 3.8，我们就发现如果 $P^n$ 是周期性地出现 0 的常返链就会造成 $P^n$ 极限不存在. 事实上，在一个正常返互通类中，若对充分大的 $n$，$\exists\, i$，$p_{ii}(n)$ 都非零那么就可以证明，全部 $p_{ij}(n)$ 当 $n$ 充分大以后也必都非零，而且这时 $P^n \to 1\pi$ ($n \to \infty$) 仍然成立. 显然，这个极限应与上段的平均极限一致. 另一方面，若不是上述情况，就必定存在一个整数 $d > 1$，使得当 $n$ 充分大时，$p_{ii}(n) > 0$ 当且仅当 $n$ 是 $d$ 的整数倍，而 $\lim_{n\to\infty} p_{ii}(nd) = d\pi_i$. 我们这里略去证明的细节.

#### 5. $f_{ij}^* \triangleq \sum_{n=1}^{\infty} f_{ij}(n)$ 的求法

注意到对 $i \neq j$

$$\begin{aligned} f_{ij}^* &= f_{ij}(1) + \sum_{n=1}^{\infty} f_{ij}(n+1) = p_{ij} + \sum_{n=1}^{\infty}\sum_{l\neq j} p_{il} f_{lj}(n) \\ &= p_{ij} + \sum_{l\neq j} p_{il} f_{lj}^*, \end{aligned}$$

可见

$$f_{ij}^* - \sum_{l\neq j} p_{il} f_{lj}^* = p_{ij} \quad (i \neq j).$$

若此方程组有唯一介于 0,1 之间的解,则此解就是 $f_{ij}^*$.

### 6. Wright-Fisher 基因频率演化问题

假设有 $a$ 与 $b$ 两种基因(例如 $a$ 是某种崎变基因,$b$ 为正常基因). 设共有 $2N$ 个个体,每个个体的基因按 $2N$ 个体种中含 $a$ 的比例为概率在下一刻成为 $a$,否则成为 $b$. 问一代代传下去两种基因个数会有何种变化趋势?

令 $\xi_n$ 为在时刻 $n$ 含基因 $a$ 的个体数,则 $\xi \triangleq \{\xi_n\}$ 是一马氏链. 因为

$$P(\xi_{n+1} = k \mid \xi_n = i, \xi_{n-1} = i_{n-1}, \cdots, \xi_0 = i_0)$$
$$= \binom{2N}{k} p^k (1-p)^{2N-k} \triangleq p_{i,k} \quad (i, k = 0, 1, \cdots, 2N),$$

其中 $p = i/2N$. 从而 $\xi$ 的转移阵为 $(p_{i,j})$. 在所有的状态中,0 与 $2N$ 是两个吸收态,即

$$P(\xi_{n+1} = k \mid \xi_n = 0) = \delta_{0,k}, \quad P(\xi_{n+1} = k \mid \xi_n = 2N) = \delta_{2N,k}.$$

而其他 $i$ 都有正概率到每个状态. 可见:

$$i \to 0, \; i \to 2N \; (0 < i < 2N),$$

但 $0, 2N$ 不能达 $i (0 < i < 2N)$,由命题 3.3,$i$ 非常返. 从而 $\lim\limits_{n \to \infty} p_{ij}(n) = 0 (0 < j < 2N)$. 为求 $L_{i,0}$ 与 $L_{i,2N}$,我们注意 $\boldsymbol{PL} = \boldsymbol{L}$,即

$$L_{i,0} = \sum_{k=0}^{2N-1} p_{ik} L_{k,0}, \quad L_{i,2N} = \sum_{k=1}^{2N} p_{ik} L_{k,2N},$$

并且

$$L_{i,0} + L_{i,2N} = 1.$$

由此不难解得

$$L_{i,0} = \left(\frac{2N - i}{2N}\right), \quad L_{i,2N} = \frac{i}{2N} \quad (i = 0, 1, \cdots, 2N).$$

回到原来提出的问题,其结论是:当经过足够多代后,基因将变为或全部是 $a$,或全部是 $b$. 若从一开始有 $i$ 个基因 $a$,则全变为 $a$ 的概率是 $i/2N$,而全变为 $b$ 的概率是 $1 - i/2N$.

## § 3.4 停时、强马氏性与马氏链的强大数律

类似于独立同分布的随机序列的强大数律，对马氏链我们也有类似的强大数律. 为了得到马氏链的强大数律，我们先引出停时及强马氏性的两个重要概念. 事实上，这两个概念本身在马氏过程的研究中都有各自独立的重要意义.

### 1. 停时

在马氏链 $\xi$ 的研究中，会出现许多随机的时间，例如前面讲到的 $\xi$ 首次到达 $j$ 的时间：

$$T_j(\omega) = \inf\{n; \xi_n(\omega) = j\},$$

它是一个随机变量，而且这个时间和 $\xi$ 的发展是"同步的"：也就是说 $\{\omega: T_j(\omega) \leqslant n\}$ 发生与否，是由 $\xi_0(\omega), \xi_1(\omega), \cdots, \xi_n(\omega)$ 来决定的：

$\{\omega: T_j(\omega) \leqslant n\} = \{\omega: \xi_0(\omega), \cdots, \xi_n(\omega)$ 中至少有一个等于 $j\}$.

又例如 $\xi(\omega)$ 首次到达某一个状态空间的子集合 $A$ 的时刻 $\tau_A(\omega)$ 也是这样的时间：

$$\{\omega: \tau_A(\omega) \leqslant n\} = \{\omega: \xi_0(\omega), \cdots, \xi_n(\omega)$$

中至少有一个在 $A$ 中$\}$. 综上分析，我们给出停时的以下定义：

**定义 3.6（停时）** 一个非负值（可取 $+\infty$）的随机变量 $\tau(\omega)$ 称为 $\xi$ 的一个停时（也有的书上将它称为马氏时间），如果 $\{\omega: \tau(\omega) \leqslant n\}$ 发生与否完全由 $\{\xi_0(\omega), \cdots, \xi_n(\omega)\}$ 决定.

马氏链 $\xi$ 首达某一状态的时刻，首达某一集合的时刻，$(\Omega, \mathscr{F}, P)$ 上两个马氏链 $\xi$ 与 $\eta$ 的相遇时间：

$$\tau(\omega) \triangleq \inf\{t; \xi_t = \eta_t\}$$

等都是停时. 后者可理解为 $\zeta \triangleq \{(\xi_t, \eta_t); t = 0, 1, 2, \cdots\}$ 的停时：

$$\{\omega: \tau(\omega) = n\} = \{\omega: \xi_m(\omega) \neq \eta_m(\omega), \forall\, m < n, \xi_n(\omega) = \eta_n(\omega)\}.$$

下面我们来考察 $\bar{T}_i(\omega) = \inf\{n \geqslant 1, \xi_n(\omega) = i\}$ 在 $\xi_0(\omega) = i$

的条件下的条件期望.事实上,

$$E_i T_i(\omega) = \sum_{k=1}^{\infty} k P_i(\omega : T_i(\omega) = k) \quad (参见(3.11)与(3.11)')$$

$$= \sum_{k=1}^{\infty} k f_{ii}(k) = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} f_{ii}(k). \tag{3.24}$$

**2. 强马氏性与强再生性**

一个时齐马氏过程从任何时间 $n_0 \geqslant 0$ 算起,仍然是一个具有同样转移矩阵的马氏链.就是说,若 $\xi$ 是一个时齐马氏链,则 $\tilde{\xi} \triangleq \{\xi_{t_0+t}; t = 0,1,2,\cdots\}$ 仍是一个马氏链,而且它们的转移矩阵相同.进一步的问题是:若在上面 $\tilde{\xi}$ 的定义中将 $t_0$ 换成一个停时(它是随机的),是否 $\tilde{\xi}$ 仍然是马氏链呢?对于离散时间参数的马氏链与马氏过程答案是肯定的(但对于后面的连续时间参数的马氏链此结论并不永远成立).下面我们对离散参数马氏链给出强马氏性定理.

**定理 3.16 (强马氏性)** 若 $\xi$ 是一个时齐离散参数马氏链,则 $\tilde{\xi}$ 也是具有同样转移概率矩阵的马氏链,其中

$$\tilde{\xi} \triangleq \{\xi_{\tau(\omega)+n}; n = 0,1,\cdots\},$$

而 $\tau(\omega)$ 是一个对 $\xi$ 的停时,且 $P(\tau(\omega) < +\infty) = 1$.

**证明** 直接验证 $\tilde{\xi}$ 满足马氏链的定义,即考察

$$P(\xi_{\tau(\omega)+n+1} = j \mid \xi_{\tau(\omega)+n} = i, \xi_{\tau(\omega)+n-1} = i_{n-1}, \cdots, \xi_{\tau(\omega)} = i_0)$$

$$= \frac{P(\xi_{\tau(\omega)+n+1} = j, \xi_{\tau(\omega)+n} = i, \xi_{\tau(\omega)+n-1} = i_{n-1}, \cdots, \xi_{\tau(\omega)} = i_0)}{P(\xi_{\tau(\omega)+n} = i, \xi_{\tau(\omega)+n-1} = i_{n-1}, \cdots, \xi_{\tau(\omega)} = i_0)}. \tag{3.25}$$

上式分子为

$$\sum_{k=0}^{\infty} P(\tau(\omega) = k, \xi_{k+n+1}(\omega) = j, \xi_{k+n}(\omega) = i,$$

$$\xi_{k-n+1}(\omega)=i_{n-1},\cdots,\xi_k(\omega)=i_0). \qquad (3.26)$$

由于 $\{\omega:\tau(\omega)=k\}=\{\omega:\tau(\omega)\leqslant k\}-\{\omega:\tau(\omega)\leqslant k-1\}$，它应由 $\xi_k(\omega),\xi_{k-1}(\omega),\cdots,\xi_0(\omega)$ 完全决定，也就是说它是使这些随机变量满足某些条件的那些 $\omega$ 组成. 所以

$$P(\omega:\tau(\omega)=k,\xi_{k+n+1}(\omega)=j,\xi_{k+n}(\omega)=i,\xi_{k+n-1}(\omega)=i_{n-1},\cdots,\xi_k(\omega)=i_0)$$
$$=P(\omega:\tau(\omega)=k,\xi_k(\omega)=i_0)p_{i_0i_1}\cdots p_{i_{n-1}i}p_{ij}.$$

同理

$$P(\omega:\tau(\omega)=k,\xi_{k+n}(\omega)=i,\xi_{k+n-1}(\omega)=i_{n-1},\cdots,\xi_k(\omega)=i_0)$$
$$=P(\omega:\tau(\omega)=k,\xi_k(\omega)=i_0)p_{i_0i_1}\cdots p_{i_{n-1}i}.$$

于是(3.26)就变成

$$\sum_{k=0}^{\infty}P(\omega:\tau(\omega)=k,\xi_k(\omega)=i_0)p_{i_0i_1}\cdots p_{i_{n-1}i}p_{ij}$$
$$=P(\omega:\xi_{\tau(\omega)}(\omega)=i_0)p_{i_0i_1}\cdots p_{i_{n-1}i}p_{ij}.$$

同样 (3.25) 的分母也应为

$$P(\omega:\xi_{\tau(\omega)}(\omega)=i_0)p_{i_0i_1}\cdots p_{i_{n-1}i}.$$

从而 (3.25) 等于 $p_{ij}$. 定理得证.

类似地，我们还可以证明以下强马氏定理：

**定理 3.17（强马氏性）** 在定理 3.16 的条件下，若 $\tau(\omega)$ 是一个停时且 $P(\tau(\omega)<\infty)=1$，$A$ 与 $B$ 分别是由 $\xi(\omega)$在 $\tau(\omega)$之后与 $\tau(\omega)$之前决定的两个事件，则

$$P(A|\xi_{\tau(\omega)}=i,B)=P(A|\xi_{\tau(\omega)}=i).$$

特别当 $A=\{\omega:\xi_{\tau(\omega)+m_1}=j_1,\cdots,\xi_{\tau(\omega)+m_r}=j_r\}$就有

$$P(\omega:\xi_{\tau(\omega)+m_1}=j_1,\cdots,\xi_{\tau(\omega)+m_r}=j_r|\xi_{\tau(\omega)}=i,B)$$
$$=P(\omega:\xi_{m_1}=j_1,\cdots,\xi_{m_r}=j_r|\xi_0=i).$$

注：对停时 $\tau(\omega)$类似于命题 3.2 的结论也成立.

### 3. 第 k 次到达 $i$ 的时刻 $t_k(\omega)$

设 $i$ 是常返态，而 $t_k(\omega)$表示 $\xi(\omega)$第 $k$ 次到达 $i$ 的时刻，即令

$$t_1(\omega) = \inf\{m > 0: \xi_m(\omega) = i\},$$
$$t_{n+1}(w) = \inf\{m > t_n: \xi_m(\omega) = i\}\ (n = 1,2,\cdots).$$
由于$\{\omega: t_n(\omega)=k\} = \{\omega: \xi_k(\omega)=i; \xi_1(\omega),\xi_2(\omega),\xi_k(\omega)$ 中恰有 $n$ 次为 $i$,即 $\xi(\omega)=i$ 发生 $n$ 次$\}$,可见 $t_n(\omega)(n\geq 1)$ 都是停时. 又由强马氏性,我们可以得到 $\{\tau_1(\omega),\tau_2(\omega),\cdots,\tau_k(\omega),\cdots\}$(其中 $\tau_k(\omega)\triangleq t_{k+1}(\omega)-t_k(\omega),k\geq1$))是一个 i.i.d. 的随机变量列. 这是因为注意到 $\xi_{t_n(\omega)}(\omega)\equiv i$,我们有

$$P(\omega:\tau_n(\omega)=k\mid\tau_{n-1}(\omega)=s_{n-1},\cdots,\tau_1=s_1)$$
$$=P(\omega:\xi_{t_n(\omega)+m}\neq i(0<m<k);\xi_{t_n(\omega)+k}=i\mid\xi_{t_n(\omega)}=i,B)$$
$$=P(\omega:\xi_m\neq i(0<m<k),\cdots,\xi_k=i\mid\xi_0=i_0)\ \text{(强马氏性)}$$
$$=P(\omega:\tau_1(\omega)=k\mid\xi_0=i),$$
其中 $B=\{\tau_{n-1}(\omega)=s_{n-1},\cdots,\tau_1=s_1\}$是 $t_n(\omega)$前发生的事件. 因上式等号右边与 $\tau_1,\cdots,\tau_{n-1}$ 取值无关,从而 $\tau_n(\omega)$ 与 $\{\tau_1,\tau_2,\cdots,\tau_{n-1}(\omega)\}$ 相互独立同分布. 而且 $P_i(\omega:\tau_1=k)=f_{ii}(k)$.

由独立同分布序列的强大数定律有:
$$\frac{1}{m-1}\sum_{k=1}^{m-1}\tau_k(\omega)\to E\tau_k(\omega)\ (m\to+\infty),$$
其中
$$E\tau_k(\omega)=E(\tau_1(\omega)\mid\xi_0=i)=\sum_k kf_{ii}(k)\triangleq\mu_i.$$
另一方面,若令 $N_n(\omega)$=在$(\xi_1(\omega),\cdots,\xi_n(\omega))$中等于 $i$ 的次数,则
$$E_i\frac{N_n(\omega)}{n}=E\left(\frac1n\sum_{k=1}^n 1_{\{i\}}(\xi_k(\omega))\mid\xi_0=i\right)$$
$$=\frac1n\sum_{k=1}^nE(1_{\{i\}}(\xi_k(\omega))\mid\xi_0=i)=\frac1n\sum_{k=1}^np_{ii}(k).$$
由定理 3.12 有:
$$\lim_{n\to\infty}E_i\frac{N_n(\omega)}{n}=\lim_{n\to\infty}\frac1n\sum_{k=1}^np_{ii}(k)=L_{ii}.$$
由于 $t_n(\omega)-t_1(\omega)=\sum_{k=1}^{n-1}\tau_k(\omega)$ 及 $N_n(\omega)$ 的定义,我们有

$$t_1 + \sum_{k=1}^{N_n(\omega)-1} \tau_k(\omega) \leqslant n < t_1 + \sum_{k=1}^{N_n(\omega)} \tau_k(\omega),$$

从而当 $N_n(\omega) \to \infty(n \to +\infty)$ 时* 就有概率为 1 地

$$\mu_i = \lim_{n\to\infty} \frac{1}{N_n(\omega)} \sum_{k=1}^{N_n(\omega)} \tau_k(\omega) = \lim_{n\to+\infty} \frac{n}{N_n(\omega)}.$$

所以(见定理 3.18 后面的注)

$$E\left(\frac{N_n(\omega)}{n}\right) \to \frac{1}{\mu_i} \quad (n \to +\infty).$$

而条件 $N_n(\omega) \to +\infty(n \to +\infty)$ 在 $i$ 常返时一定成立.

*注:事实上,因为其余集的概率为

$$\begin{aligned}
P(\{N_n(\omega) \to +\infty\}^c) &= P\left(\bigcup_{m=1}^{\infty} \bigcap_{n=1}^{\infty} \{N_n(\omega) < m\}\right) \\
&\leqslant \sum_{m=1}^{\infty} P\left(\bigcap_{n=1}^{\infty} \{N_n(\omega) < m\}\right) \quad (\text{注意 } N_n \leqslant N_{n+1}) \\
&= \sum_{m=1}^{\infty} \lim_{n\to\infty} P(N_n(\omega) < m) \\
&= \sum_{m=1}^{\infty} \lim_{n\to\infty} P(t_m > n) \\
&= 0 \quad \left(\text{注意 } t_m = t_1 + \sum_{k=1}^{m-1} \tau_k \to +\infty\right).
\end{aligned}$$

综上所述,我们得到下面进一步刻画 $L$ 的定理.

**定理 3.18** 当 $i$ 是一个常返状态时,则概率为 1 地有

$$\lim_{n\to\infty} \frac{1}{n} N_n(\omega) = \mu_i^{-1}, \quad L_{ii} = \frac{1}{\mu_i},$$

其中 $\mu_i = E(T_i(\omega) | \xi_0(\omega) = i) = \sum_{k=1}^{\infty} k f_{ii}(k)$. 当 $i$ 是非常返状态时, $\lim_{n\to\infty} p_{ii}(n) = 0$, 因而也有

$$L_{ii} = 0.$$

**推论 1** 设 $f$ 是状态空间 $S$ 上的一个实函数, $\xi$ 是有限状态不可约常返马氏链(即只有一个常返类),则概率为 1 地

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n}f(\xi_k)=\sum_{i\in S}\frac{f(i)}{\mu_i}.$$

**证明** 由于

$$f(\xi_k)=\sum_{i\in S}f(i)I_{\{i\}}(\xi_k),$$

我们立即由定理得到

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n}f(\xi_k)&=\frac{1}{n}\sum_{k=1}^{n}\sum_{i\in S}f(i)I_{\{i\}}(\xi_k)\\&=\sum_{i\in S}f(i)\frac{1}{n}\sum_{k=1}^{n}I_{\{i\}}(\xi_k)\\&=\sum_{i\in S}f(i)\left\{\frac{1}{n}N_n(\omega)\right\}\to\sum_{i\in S}\frac{f(i)}{\mu_i}.\end{aligned}$$

**推论 2** 设 $\xi$ 不可约，常返，则对任意状态 $i$ 与 $j$，

$$L_{ji}=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}p_{ji}(k)=\frac{1}{\mu_i}.$$

可见 $L$ 是每行都相同的矩阵，而且 $L_{ji}$ 或全为零，或全不为零.

**证示** 由于 $\frac{1}{n}N_n(\omega)\to 1/\mu_i,\mu_i>0$ 可见*

$$E_j\left\{\frac{1}{n}N_n(\omega)\right\}\to 1/\mu_i\quad(n\to+\infty).$$

但前面定理的证明中已得到

$$\begin{aligned}E_j\frac{1}{n}N_n(\omega)&=\frac{1}{n}\sum_{k=1}^{n}E_j(I_{\{i\}}(\xi_k))=\frac{1}{n}\sum_{k=1}^{n}P_j(\xi_k(\omega)=i)\\&=\frac{1}{n}\sum_{k=1}^{n}p_{ji}(k)\to L_{ji},\end{aligned}$$

于是* $L_{ji}=1/\mu_i$.

* 注:此两处我们事实上承认了

$$\lim_{n\to\infty}E_j\left\{\frac{1}{n}N_n(\omega)\right\}=\lim_{n\to\infty}\left\{\frac{1}{n}E_jN_n(\omega)\right\}.$$

对状态空间是可数的情形，在数学上要严格说明这一步的正确性要用到测度论(详见 § 3.7 段).

另一方面由于 $\xi$ 的各状态互通，对任意二状态 $i$ 与 $j$，都存在

正整数 $M$ 与 $N$ 使得

$$p_{ij}(M)>0,\quad p_{ji}(N)>0.$$

于是 $p_{ii}(M+k+N)\geqslant p_{ij}(M)p_{jj}(k)p_{ji}(N)$. 由此得到

$$\begin{aligned}
L_{ii}&=\lim_{n\to+\infty}\frac{1}{n}\sum_{k=1}^{n}p_{ii}(k)=\lim_{n\to+\infty}\frac{1}{n+M+N}\sum_{k=1}^{n+M+N}p_{ii}(k)\\
&=\lim_{n\to+\infty}\frac{1}{n}\sum_{k=1}^{n}p_{ii}(M+N+k)\\
&\geqslant\lim_{n\to+\infty}\frac{1}{n}\sum_{k=1}^{n}p_{jj}(k)p_{ij}(M)p_{ji}(N)\\
&=L_{jj}p_{ij}(M)p_{ji}(N),
\end{aligned}$$

即"$L_{jj}>0$"蕴含"$L_{ii}>0$". 可见 $S$ 中有一个状态 $j$ 使 $L_{jj}>0$,就必有全部 $L_{kk}=L_{ii}>0$.

**定义 3.7** 对常返状态 $i$,若有 $L_{ii}=0$,即 $\mu_i=\sum_k kf_{ii}(k)=+\infty$,则称 $i$ 为**零常返状态**;若 $L_{ii}>0$,即 $\mu_i<+\infty$,则称 $i$ **为正常返状态**.

推论 2 说明一个互通常返类中的状态或全部是正常返,或全部是零常返. 当 $\xi$ 是互通零常返时, $\lim_{n\to+\infty}\frac{1}{n}\sum_{k=1}^{n}p_{ij}(k)=\pi_j=0$ 而得不到一个不变分布 $\pi=(\pi_j;j\in S)$,因为这时全部 $\pi_j$ 为零,不成其为分布了. 其实这时根本不存在不变(概率)分布.

推论 1 对于马氏链的概率特征的统计具有十分重要的意义,它告诉我们一个马氏链的函数的数学期望等数值特征都可以用此马氏链在一条轨道上的现实来估计. 下面的推论 3 给出了 $(p_{ij})$, $\pi_i$ 的强相合估计量(请参考数理统计中强相合估计的定义):

**推论 3** 对于不可约正常返马氏链,我们以概率为 1 地有:

$$\pi_i p_{ij}=\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}1_{\{i\}}(\xi_k(\omega))1_{\{j\}}(\xi_{k+1}(\omega)),$$

$$\pi_i=\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}1_{\{i\}}(\xi_k(\omega)),$$

$$p_{ij} = \lim_{n\to\infty} \frac{\sum_{k=0}^{n-1} 1_{\{i\}}(\xi_k(\omega)) 1_{\{j\}}(\xi_{k+1}(\omega))}{\sum_{k=0}^{n-1} 1_{\{i\}}(\xi_k(\omega))}.$$

**证明** 由于 $N_n(\omega) = \sum_{k=0}^{n} 1_{\{i\}}(\xi_k(\omega))$，因而本推论的第二式是定理3.18及推论2的直接结果. 为了得到第一个等式，考虑状态空间$\mathscr{S}' = \{(i,j): p_{ij} > 0\}$上的如下马氏链

$$\eta_n(\omega) = (\xi_n(\omega), \xi_{n+1}(\omega)),$$

于是由

$$\begin{aligned} & P(\eta_{n+1}(\omega) = (j_1, j_2) | \eta_n(\omega) = (i_1, i_2), \eta_{n-1}(\omega) = (i_{n-1}, i'_n), \cdots, \\ & \qquad \eta_0(\omega) = (i_0, i'_1)) \\ & = P(\xi_{n+1}(\omega) = j_1, \xi_{n+2}(\omega) = j_2 | \xi_{n+1}(\omega) = i_2, \xi_n(\omega) = i_1, \cdots, \\ & \qquad \xi_1(\omega) = i'_1, \xi_0(\omega) = i_0) \\ & = P(\xi_{n+1}(\omega) = j_1, \xi_{n+2}(\omega) = j_2 | \xi_{n+1}(\omega) = i_2) = p_{j_1 j_2} \delta_{i_2 j_1}. \end{aligned}$$

可见 $\{\eta_n(\omega): n = 0, 1, \cdots\}$ 也是一个马氏链，可以验证 $\{\eta_n\}$ 仍是互通的、常返的，而且由 $\{\pi_i\}$ 是 $\xi$ 的不变分布可得

$$\sum_{i_1, i_2} \pi_{i_1} p_{i_1 i_2} p_{j_1 j_2} \delta_{i_2 j_1} = \sum_{i_1} \pi_{i_1} p_{i_1 j_1} p_{j_1 j_2} = \pi_{j_1} p_{j_1 j_2}.$$

再将定理 3.18 与推论 2 用到 $\{\eta_n\}$ 上，就得到

$$\begin{aligned} & \frac{1}{n} \sum_{k=0}^{n-1} I_{\{i\}}(\xi_k(\omega)) I_{\{j\}}(\xi_{k+1}(\omega)) \\ & \quad = \frac{1}{n} \sum_{k=0}^{n-1} 1_{\{(i,j)\}}(\xi_k(\omega), \xi_{k+1}(\omega)) \to \pi_i p_{ij}, \quad 当 n \to +\infty. \end{aligned}$$

最后，第三式是第一、二式的直接推论. 定理证毕.

若将 $\{\pi_i\}$ 取为不可约正常返马氏链的初始分布，我们就得到 $\{\xi_n, n \geq 0\}$ 是一个平稳过程，而且对任意有界函数 $f$ 概率为 1 地有

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(\xi_k(\omega)) = \sum_j \pi_j f(j) = Ef(\xi_1).$$

因而有时人们也称不可约正常返马氏链为**遍历链***.

* 注:请参见第七章遍历定理.

## §3.5 禁忌概率、首出时、首中时与首中分布

### 1. 概念及关系

考虑以 $\mathscr{S}$ 为状态空间的马氏链 $\{\xi_n\}$. 设 $A$ 是 $\mathscr{S}$ 的一个子集.

**定义 3.8(禁忌概率)** 由 $i$ 出发,经 $m$ 步到达 $j$,且禁忌进入 $A$ 的概率指

$$P_i(\omega;\text{对 } 0 < m < n \text{ 都有 } \xi_m(\omega) \notin A, \xi_n(\omega) = j),$$

以后我们把它记为 ${}_Ap_{ij}(n)$.

上面的定义意思是 ${}_Ap_{ij}(n)$ 表示从 $i$ 出发的条件下,经 $n$ 步 $\xi$ 达 $j$, 但是在此之前从未到过 $A$ 的概率.

**定义 3.9 (首中时与首中分布)** 我们将 $\xi$ 除时刻 0 以外首次到达 $A$ 的时间

$$\tau_A(\omega) = \inf\{n > 0;\xi_n(\omega) \in A\}$$

称为 $\xi$ 对 **$A$ 的首中时**;而把

$$P_i(\omega;\tau_A(\omega) < +\infty, \xi_{\tau_A(\omega)} = j)$$

称为从 $i$ 出发,$\xi$ **首中 $A$ 的分布**(当 $P_i(\tau_A(\omega) < \infty) = 1$ 时,首中 $A$ 的分布是概率分布,否则 $\sum\limits_{j\in A} P_i(\tau_A < \infty, \xi_{\tau_A} = j)$ 可能小于 1).

又令

$$\sigma_A(\omega) = \inf\{n > 0;\xi_n(\omega) \notin A\},$$

$\sigma_A(\omega)$ 称为 $\xi$ **对 $A$ 的首出时**. 事实上,令 $B = A^c = \mathscr{S} - A$, 则我们有

$$\sigma_A(\omega) = \tau_B(\omega).$$

不难看出,以上概念之间有以下关系:

**命题 3.19** 1) ${}_Ap_{ij}(n) = P_i(\tau_A(\omega) \geqslant n, \xi_n(\omega) = j)$. 又当 $j \notin A$ 时我们有

$${}_A p_{ij}(n) = P_i(\tau_A(\omega) > n, \xi_n(\omega) = j);$$

而当 $j \in A$ 时有

$${}_A p_{ij}(n) = P_i(\tau_A(\omega) = n, \xi_n(\omega) = j).$$

2) $${}_A p_{ij}(n) = P_i(\xi_m(\omega) \in B, \forall\, m < n, \xi_n(\omega) = j)$$
$$= (P_B^n)_{ij} (\forall\, i, j \notin A, B = A^c),$$

其中 $P_B = (p_{ij})_{i,j \in B}$;

3) $$P_i(\tau_A(\omega) > n) = (P_B^n \mathbf{1})_i = \sum_{j \in B} {}_A p_{ij}(n) (\forall\, i \in B = A^c);$$

4) 当 $P_i(\tau_A(\omega) < \infty) = 1$ 时有

$$E_i \tau_A(\omega) = \left( \sum_{n=0}^{\infty} P_B^n \mathbf{1} \right)_i$$

(上式蕴含等式两边同时有限或无限).

**证明** 1°) 在 1)中的各论断是定义的直接推论.

2°) 当 $n = 1$ 时有 ${}_A p_{ij}(1) = p_{ij}$. 下面我们用归纳法. 设对 $i, j \notin A$ 等式 ${}_A p_{ij}(n) = (P_B^n)_{ij}$ 正确. 于是

$${}_A p_{ij}(n+1) = P_i(\omega : \tau_A(\omega) \geqslant n+1, \xi_{n+1}(\omega) = j)$$
$$= \sum_{l \notin A} P_i(\omega : \tau_A(\omega) \geqslant n, \xi_n(\omega) = l, \xi_{n+1}(\omega) = j)$$
$$= \sum_{l \in B} {}_A p_{il}(n) P(\omega : \xi_{n+1}(\omega) = j \mid \xi_n(\omega) = l)$$
$$= \sum_{l \in B} (P_B^n)_{il}\, p_{lj} = (P_B^{n+1})_{ij}.$$

由归纳法 2)得证.

3°) $$P_i(\tau_A(\omega) > n) = P_i(\omega : \tau_A(\omega) \geqslant n, \xi_n(\omega) \notin A)$$
$$= \sum_{j \notin A} P_i(\omega : \tau_A(\omega) \geqslant n, \xi_n(\omega) = j)$$
$$= \sum_{j \in B} {}_A p_{ij}(n) = \sum_{j \in B} (P_B^n)_{ij}$$
$$= (P_B^n \mathbf{1})_i$$ ($\mathbf{1}$ 为分量全为 1 的列向量).

4°) 当 $P_i(\tau_A(\omega) < \infty) = 1$ 时,我们有

$$E_i \tau_A(\omega) = \sum_{k=1}^{\infty} k P_i(\omega : \tau_A(\omega) = k) = \sum_{k=1}^{\infty} \sum_{n=1}^{k} P_i(\omega : \tau_A(\omega) = k)$$

$$= \sum_{n=1}^{\infty}\sum_{k=n}^{\infty} P_i(\omega:\tau_A(\omega)=k) = \sum_{n=1}^{\infty} P_i(\omega:\tau_A(\omega)\geqslant n)$$

$$= \sum_{n=0}^{\infty} P_i(\omega:\tau_A(\omega)> n) = \left(\sum_{n=0}^{\infty} P_B^n \mathbf{1}\right)_i.$$

**命题 3.20** 若状态集 $C$ 满足：$\forall\, i\in B$ 均有 $P_i(\tau_C(\omega)<\infty)=1$，且 $B\cap C=\varnothing$，则对 $\forall\, i,j\in B$ 有

$$\sum_{n=0}^{\infty} P_B^n < \infty \tag{3.26}$$

及

$$(I-P_B)^{-1} = \sum_{n=0}^{\infty} P_B^n. \tag{3.27}$$

再则，如果对于 $\forall\, i\in B$ 均有

$$p_i(\tau_{B^c}(\omega)<\infty) = 1, \tag{3.28}$$

那么我们有

$$E_i\tau_{B^c}(\omega) = ((I-P_B)^{-1}\mathbf{1})_i. \tag{3.29}$$

$$(P_i(\omega:\xi_{\tau_{B^c}}(\omega)=j))_{i\in B} = (I-P_B)^{-1}(p_{kj})_{k\in B}. \tag{3.30}$$

**证明** 1) 令 $D\triangleq B^c\supset C$.

$$\widetilde{p}_{kj} = \begin{cases} p_{kj}, & k\notin D; \\ \delta_{kj}, & k\in D, \end{cases}$$

$$\widetilde{P} = (\widetilde{p}_{kj}).$$

显见用 $\widetilde{P}$ 代替 $P$ 时，相应于命题的条件仍然成立. 然而 $D$ 在 $\widetilde{P}$ 下不可能到达 $D$ 以外任何状态. 可见在 $\widetilde{P}$ 下除 $D$ 外都不是常返态，故而

$$\sum_{n}\widetilde{p}_{ij}(n) < \infty.$$

再则，对 $n$ 充分大我们有

$$(P_B^n)_{ij} = {}_D p_{ij}(n) = \widetilde{p}_{ij}(n),$$

所以(3.26)成立.

另一方面，我们有 $\left(\sum_{n=0}^{\infty} P_B^n\right)P_B = \sum_{n=1}^{\infty} P_B^n = \left(\sum_{n=0}^{\infty} P_B^n - I\right)$. 也

就是

$$\left(\sum_{n=0}^{\infty}P_B^n\right)(\boldsymbol{I}-P_B)=\boldsymbol{I}.$$

类似地

$$(\boldsymbol{I}-P_B)\left(\sum_{n=0}^{\infty}P_B^n\right)=\boldsymbol{I}.$$

从而得到(3.27)．再利用定理 3.19,4)及 2)和马氏性便得到在(3.28),(3.29)及(3.30)的正确性.

**推论** 若状态集 $\mathscr{S}$ 为有限集合且有

$$P_i(\omega:\tau_A(\omega)<+\infty)=1\quad(\forall\, i\in A^c),$$

则

$$E_i\tau_A(\omega)=\sum_{k\in A^c}D_{ik}(A^c)/D(A^c),\qquad(3.31)$$

及对 $j\in A$ 有

$$P_i(\omega:\xi_{\tau_A}(\omega)=j)=-D_{ij}(A^c\cup\{j\})/D(A^c),\quad(3.32)$$

其中 $D(B)$表示$(\boldsymbol{I}-P_B)$的行列式,$D_{jk}$表示$(\boldsymbol{I}-P_B)$中位置在第 $j$ 行第 $k$ 列的元的代数余子式.

**证明** 对于 $B=A^c$ ,由命题 3.20 及逆矩阵的代数余子式表示立得(3.31). 又(3.30)左方 是方程

$$(\boldsymbol{I}-P_B)\vec{x}=(p_{ij})_{i\in B}\quad(B=A^c)$$

的唯一解,利用克莱姆法则便得到(3.32).

## 2. 蛋白质演化的马氏链模型

本节问题选自 R. N. Curnow 的论文:"The use of Markov chain models in studying the evolution of the proteins ",J. Theor. Biol. (1988)134,51--7.

下面我们用前面讲过的马氏链的知识来解决这一问题.

蛋白质是由许多氨基酸构成的. 全部氨基酸共有 20 种,它们分别是由选自 ($U$, $C$,$A$,$G$) 4 种核甘酸的三元组编码:$UUU$,

$UUC, UUA, UUG, UCU, \cdots, GGG$（共 64 种）组成（详见表 3.1）. 在蛋白质变异过程中，编码发生如下变化：每次变异只允许有一个核甘酸变化为其他 3 种核甘酸，于是从每个确定的编码出发，变化后的可能结果共有 9 种，例如 $AAA$ 可以变为 $AAC$，$AAG$，$AAU$，$ACA$，$AGA$，$AUA$，$CAA$，$GAA$，$UAA$ 9 种. 又设各种变化是平等的，于是 $AAA$ 就各以 $\frac{1}{9}$ 的概率分别变为上述 9 种之一. 于是氨基酸编码的变异就是一个马氏链 $\{\xi_n\}$，其中 $\xi_n$ 表示在时刻 $n$ 的氨基酸的编码. 若按字典排列法（字母以 $U$，$C$，$A$，$G$ 为序）将 64 种氨基酸编码顺序排列，则 $S = \{UUU, UUC, \cdots, GGG\}$ 是其状态空间. 其转移概率阵如下：

$$\boldsymbol{P} = \frac{1}{9}\begin{pmatrix} \boldsymbol{E} & \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{E} & \boldsymbol{I} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{E} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{E} \end{pmatrix},$$

其中 $\boldsymbol{I}$ 为 $16 \times 16$ 单位阵，

$$\boldsymbol{E} = \begin{pmatrix} \boldsymbol{F} & \boldsymbol{I}_4 & \boldsymbol{I}_4 & \boldsymbol{I}_4 \\ \boldsymbol{I}_4 & \boldsymbol{F} & \boldsymbol{I}_4 & \boldsymbol{I}_4 \\ \boldsymbol{I}_4 & \boldsymbol{I}_4 & \boldsymbol{F} & \boldsymbol{I}_4 \\ \boldsymbol{I}_4 & \boldsymbol{I}_4 & \boldsymbol{I}_4 & \boldsymbol{F} \end{pmatrix},$$

$\boldsymbol{I}_4 = 4 \times 4$ 单位阵.

$$\boldsymbol{F} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

容易看出 $\boldsymbol{P}$ 是不可约的（互通的）.

但是在 64 种编码中 $UUA, UGA, UAG$ 三种是"死状态"，即一旦演化到它们以后，氨基酸就不再演化. 其含义译为马氏链的语言就是考虑将三种"死状态"组成的集合列为禁忌集 $\Delta$ 还有 61 种编

码构成 20 个氨基酸，所以一个氨基酸中可能包含不同种的编码（见表 3）．在 R. N. Curnow 的论文所提出的，需要研究的问题是：

1）从某一氨基酸 $B_k$ 出发（设同一个氨基酸的各编码都平等）平均经过多少次变异而后“死亡”．

2）从某一氨基酸 $B_k$（它是编码组成的某个集合）出发，在“死亡”前到达氨基酸 $B_l$ 的次数．

3）设初始分布在各氨基酸是均匀的（但是由于各个氨基酸所含编码数是不相等的，所以初始分布在全体编码上并不平均），求演化过程中在死亡前到达各氨基酸的平均次数．

若 $B_k$ 表示第 $k$ 种氨基酸中所含编码的集合，设 $B_k$ 中编码的个数为 $N_k$，则将 1）～3）译为马氏链的语言，我们有下面的 1′）～3′）：

1′）求 $t_k \triangleq \dfrac{1}{N_k}\sum\limits_{\xi\in B_k}E_\xi(\tau_\Delta(\omega))$（这里 $\Delta$ 为三个“死状态”的集合：$\Delta=\{UUA,UGA,UAG\}$）．

事实上，令 $B=\Delta^c$，则由（3.29）及（3.31）

$$t_k=\frac{1}{N_k}\sum_{\xi\in B_k}\sum_{n=1}^{\infty}((\boldsymbol{P}_B^n)\mathbf{1})_\xi=\frac{1}{N_k}\sum_{\xi\in B_k}E_\xi((\boldsymbol{I}-\boldsymbol{P}_B)^{-1}\mathbf{1})_\xi$$

$$=\frac{1}{N_k}\sum_{\xi\in B_k}\sum_{\eta\in B}D_{\eta\xi}(B)/D(B).$$

其中 $D_{\eta\xi}(B)$ 是 $(\boldsymbol{I}-\boldsymbol{P}_B)$ 的 $\xi$ 行 $\eta$ 列的代数余子式，$D(B)$ 是 $(\boldsymbol{I}-\boldsymbol{P}_B)$ 的行列式．

2′）求 $\psi_{kl}\triangleq\dfrac{1}{N_k}\sum\limits_{\xi\in B_k}E_\xi\{\sum\limits_{n=1}^{\tau_\Delta(\omega)-1}1_{(B_l)}(\xi_n)\}$．

事实上，由命题 3.19 之 3°）与 4°），我们有

$$\psi_{kl}=\frac{1}{N_k}\sum_{\xi\in B_k}\sum_{\eta\in B_l}E_\xi\{\sum_{n=1}^{\tau_\Delta(\omega)-1}1_{\{\eta\}}(\xi_n)\}$$

$$=\frac{1}{N_k}\sum_{\xi\in B_k}\sum_{\eta\in B_l}\sum_{m=0}^{\infty}P_\xi(\tau_\Delta(\omega)>m,\xi_m(\omega)=\eta)$$

$$= \frac{1}{N_k}\sum_{\xi\in B_k}\sum_{\eta\in B_l}\sum_{m=0}^{\infty} {}_{\Delta}P_{\xi\eta}(m)$$

$$= \frac{1}{N_k}\sum_{\xi\in B_k}\sum_{\eta\in B_l}\frac{D_{\xi\eta}(B)}{D(B)} \quad (B = \Delta^c).$$

$$3'）\ \psi_k \triangleq \frac{1}{20}\sum_{l=1}^{20}\frac{1}{N_k}\sum_{\xi\in B_k}E_\xi\{\sum_{n=1}^{\tau_\Delta(\omega)-1}1_{\{B_l\}}(\xi_n)\} = \frac{1}{20}\sum_{k=1}^{20}\psi_{kl}.$$

值得指出的是：在生物学中，有时也考虑以氨基酸为状态，但是如果我们现在的框架是合理的话，那么以氨基酸为状态的变异过程就不是马尔可夫的（参见本节习题）.

**表 3.1 氨基酸编（密）码表**

| 名　　称 | 编　　码 | 名　　称 | 编　　码 |
|---|---|---|---|
| Phe 苯丙氨酸 | $UUU,UUC$ | His 组氨酸 | $CAU,CAC$ |
| Leu 亮氨酸 | $UUA,UUG,CUU,$<br>$CUC,CUA,CUG$ | Gln 谷氨酰胺 | $CAA,CAG$ |
| Ile 异亮氨酸 | $AAU,AUC,AUA$ | Asn 天门冬酰胺 | $AAU,AAC$ |
| Met 旦氨酸 | $AUG$ | Lys 赖氨酸 | $AAA,AAG$ |
| Val 缬氨酸 | $GUU,GUC,GUA,GUG$ | Asp 天门冬氨酸 | $GAU,GAC$ |
| Ser 丝氨酸 | $UCU,UCC,UCA,$<br>$UCG,AGU,AGC$ | Glu 谷氨酸 | $GAA,GAG$ |
| Cys 半胱氨酸 | $UGU,UGC$ | Trp 色氨酸 | $UGG$ |
| Pro 脯氨酸 | $CCU,CCC,CCA,CCG$ | Arg 精氨酸 | $CGU,CGC,CGA,$<br>$CGG,AGA,AGG$ |
| Thr 苏氨酸 | $ACU,ACC,ACA,ACG$ | Ala 丙氨酸 | $GCU,GCC,GCA,GCG$ |
| Tyr 酪氨酸 | $UAU,UAC$ | 甘氨酸 | $GGU,GGC,GGA,GGG$ |
| 终止编码 | $UAA,UAG,UGU$ | | |

## §3.6 应 用 例 题

### 1. Gibbs 样本生成法（Gibbs Sampler）

在许多很复杂的统计问题中，有时很难对各种统计方法进行理论分析．为了评估它们的优劣，常见的实用办法是作随机模拟：

即设法按问题的要求与条件去构造出一系列模拟样本，对它们试用我们提出的统计方法作统计推断，观察对这些模拟样品作出的推断的正确率. Gibbs 样本生成法就是一种有效可行的高维分布总体的样本生成法. 它的基本想法是构造一个不可约正常返的马氏链 $\xi=\{\xi_n\}$，使得这个马氏链的不变分布正好是某个给定的分布 $\pi$(即前面所提到的高维总体). 由于当 $n$ 充分大时，$\xi_n$ 的分布渐近于分布 $\pi$，所以这时 $\xi_n$ 可以近似地作为以 $\pi$ 为分布的总体的样本. 本章考虑的是可数状态(设状态空间为 $\mathscr{S}$)马氏链，所以我们这里只对 $\pi$ 是在可数集上取值的分布来讨论，但其处理问题的思想也适用于生成连续分布的随机变量的样本.

我们知道，给定一个一维分布，容易得到一个以它为分布的随机变量的样本；例如可以简单的利用 $SAS$ 软件得到. 但是给定一个 $m$ 维 $(m>1)$ 联合分布，特别是当 $m$ 较大时，事情就远非如此简单. 令 $\mathscr{S}\triangleq\mathscr{S}_1\times\cdots\times\mathscr{S}_m$，其中 $\mathscr{S}_i$ 对不同的 $i$ 是同一个一维空间. 设给了一个 $m$ 维联合分布 $\pi(x_1,x_2,\cdots,x_m)$，我们考虑如下的转移概率

$$p_{xy}\triangleq p(x,y)=\prod_{k=1}^{m}\pi(y_k|y_1,y_2,\cdots,y_{k-1},x_{k+1},\cdots,x_m)\,,\tag{3.33}$$

其中 $x=(x_1,x_2,\cdots,x_m)$，$y=(y_1,y_2,\cdots,y_m)$，$x_i,y_i\in\mathscr{S}_i$，而

$$\pi(y_k|y_1,y_2,\cdots,y_{k-1},x_{k+1},\cdots,x_n)$$

是在除第 $k$ 个分量外，将第 1 至 $k-1$ 个分量固定为 $y_1,y_2,\cdots,y_{k-1}$，并将第 $k+1$ 至第 $n$ 个分量固定为 $x_{k+1},\cdots,x_n$ 的条件下第 $k$ 个分量在 $y_k$ 处的条件分布.

我们来验证 $\pi(x_1,x_2,\cdots,x_m)$ 正是以 $(p_{x,y})$ 为转移阵(请读者自己验证 $(p_{x,y})$ 确实是一个转移阵，即 $\sum\limits_{y}p_{x,y}=1$)的马氏链的不变分布. 事实上，我们有

$$\sum_{(x_1,\cdots,x_m)} \pi(x_1,\cdots,x_m)\ p((x_1,\cdots,x_m),(y_1,\cdots,y_m))$$

$$= \sum_{x_m}\sum_{x_{m-1}}\cdots\sum_{x_1}\pi(x_1,\cdots,x_m)\pi(y_1|x_2,\cdots,x_m)$$

$$\times\ \pi(y_2|y_1,x_3,\cdots,x_m)\cdots\pi(y_m|y_1,\cdots,y_{m-1})$$

$$= \sum_{x_m}\sum_{x_{m-1}}\cdots\left[\sum_{x_1}\pi(x_1,x_2,\cdots,x_m)\ \frac{\pi(y_1,x_2,\cdots,x_m)}{\sum_x \pi(x,x_2,\cdots,x_m)}\right]$$

$$\times\ \pi(y_2|y_1,x_3,\cdots,x_m)\cdots\pi(y_m|y_1,\cdots,y_{m-1})$$

$$= \sum_{x_m}\sum_{x_{m-1}}\cdots\sum_{x_2}\pi(y_1,x_2,\cdots,x_m)$$

$$\times\ \pi(y_2|y_1,x_3,\cdots,x_m)\cdots\pi(y_m|y_1,\cdots,y_{m-1})$$

$$= \sum_{x_m}\sum_{x_{m-1}}\cdots\sum_{x_2}\pi(y_1,x_2,x_3,\cdots,x_m)$$

$$\times\ \frac{\pi(y_1,y_2,x_3,\cdots,x_m)}{\sum_x \pi(y_1,x,x_3,\cdots,x_m)}\cdots\pi(y_m|y_1,\cdots,y_{m-1}) = \cdots$$

$$= \sum_{x_m}\pi(y_1,y_2,\cdots,y_{m-1},x_m)\ \times\ \frac{\pi(y_1,y_2,\cdots,y_m)}{\sum_x \pi(y_1,y_2,\cdots,y_{m-1},x)}$$

$$= \pi(y_1,y_2,\cdots,y_m).$$

我们需要指出：

(1) 我们并不需要知道 $\pi(x_1,x_2,\cdots,x_m)$ 的表达式究竟是什么，而只需要知道在固定其他一切分量的条件下，余下的某一个分量的条件分布，就可以得到以 $(p_{x,y})$ 为转移阵的时齐马氏链 $\xi$ 的样本. 换句话说，当 $\pi(x_1,x_2,\cdots,x_m)$ 相差一个常数因子确定时，$(P_{x,y})$就完全确定了.

(2) 由马氏链 $\xi_n(\omega)$ 的样本去求得下一步 $\xi_{n+1}(\omega)$ 的样本时，由于在(3.33)式中的 $p_{x,y}$ 是一维条件分布的乘积，获得遵从这些一维条件分布的样本时，我们可以利用生成此一维分布的现成统计软件，即可按以下程序逐个得到 $\xi_{n+1}(\omega)$ 的样本的各个分量

$(y_1, y_2, \cdots, y_m)$:

先得到服从分布 $\{\pi(y_1|x_2, \cdots, x_m): y_1 \in \mathscr{S}_1\}$ 的随机变量 $\xi_{n+1,1}(\omega)$ 的一个样本 $y_1$;

再得到服从分布 $\{\pi(y_2|y_1, x_3, \cdots, x_m): y_2 \in \mathscr{S}_2\}$ 的随机变量 $\xi_{n+1,2}(\omega)$ 的一个样本 $y_2$;

依此下去,得到服从分布 $\{\pi(y_k|y_1, \cdots, y_{k-1}, x_{k+1}, \cdots, x_m): y_k \in \mathscr{S}_k\}$ 的随机变量 $\xi_{n+1,k}(\omega)$ 的一个样本 $y_k(k = 1, \cdots, m-1)$;

最后得到服从分布 $\{\pi(y_m|y_1, \cdots, y_{m-1}): y_m \in \mathscr{S}_m\}$ 的随机变量 $\xi_{n+1,m}(\omega)$ 的一个样本 $y_m$.

定义

$$y \triangleq (y_1, y_2, \cdots, y_m),$$

它就是 $\xi_{n+1}(\omega)$ 的一个样本.

现在任取 $\xi_0(\omega) = y^{(0)}$,按上面方法可以得到 $\xi_1(\omega)$ 的一个样本 $y^{(1)}, \cdots$. 对 $n$ 归纳地可用上面方法得到 $\xi_2(\omega), \cdots, \xi_n(\omega)$ 的样本 $y^{(2)}, \cdots, y^{(n)}$. 当 $n$ 充分大时,由于马氏链 $\xi_n(\omega)$ 的分布近似于 $\pi(x_1, x_2, \cdots, x_m)$,我们就可以认为 $y^{(n)}$ 就是(近似地)服从分布 $\pi(x_1, x_2, \cdots, x_m)$ 的一个样本.

因为这里用了模拟马氏链的思想,所以 Gibbs 样本生成法又称为马氏链 Monte Carlo 方法.

## 2. 可逆马氏链

正如§3.3 例 3.6 中给出的马氏链那样,一个马氏链可以在一些状态间,按照一定的网络结构转移(该例中各状态即各车站),状态之间按车站间是否有道路而规定联结.该过程在每一步的状态迁移只能沿网络规定的可能状态进行.在例 3.6 中,这种迁移就是汽车每次从它当时所在车站随机地选择一个有道路直接联结的车站前进.进而,我们还可以考虑在网络中相邻的两个状态(设为状态 $i, j$)之间有阻尼(设为 $\beta_{ij} = \beta_{ji} > 0$)的情形,并规定由状态 $i$ 迁移到状态 $j$ 的概率与阻尼成反比,即

$$p_{ij} = \alpha_i/\beta_{ij}, \quad P = (p_{ij}).$$

这里，$\alpha_i = \left( \sum_j \frac{1}{\beta_{ij}} \right)^{-1}$. 这样的马氏链具有一个重要的特点：存在 $\underset{\sim}{\mu} = (\mu_1, \cdots, \mu_N)$，使得

$$\mu_i p_{ij} = \mu_j p_{ji}. \tag{3.34}$$

这里我们可以取 $\mu_i = C/\alpha_i$, $C = \left( \sum_i 1/\alpha_i \right)^{-1}$，于是我们就有

$$\mu_i p_{ij} = C/\beta_{ij} = C/\beta_{ji} = \mu_j p_{ji}.$$

**定义 3.8（可逆马氏链）** 设马氏链 $\xi$ 以 $P$ 为转移阵，若存在分布 $\underset{\sim}{\mu} = (\mu_1, \cdots, \mu_N)$ 使得 $\mu_i p_{ij} = \mu_j p_{ji} (\forall\ i, j \in \{1, \cdots, N\})$（对 $N = +\infty$ 的情形，定义同样适用），则称 $\xi$ 为一个可逆马氏链，称之为 $\xi$ 的可逆平稳分布（或可逆不变分布）.

事实上，$\underset{\sim}{\mu}$ 是 $P$ 的一个不变分布，因为

$$\sum_i \mu_i p_{ij} = \sum_i \mu_j p_{ji} = \mu_j,$$

于是正如 §3.3 指出的：若将 $\xi$ 的初分布取为 $\underset{\sim}{\mu}$，则 $\xi$ 是一个分布按时间平移不变的随机过程（也称平稳过程）. 不仅如此，这个过程还有对时间倒逆的不变性，也就是说，如果将时间反过来看，该过程的分布不变：对 $\forall\ n \leqslant m$

$$\begin{aligned}
&P(\omega: \xi_n(\omega) = i_n, \xi_{n+1}(\omega) = i_{n+1}, \cdots, \xi_m(\omega) = i_m) \\
&\quad = \mu_{i_n} p_{i_n i_{n+1}} \cdots p_{i_{m-1} i_m} = \mu_{i_{n+1}} p_{i_{n+1} i_n} p_{i_{n+1} i_{n+2}} \cdots p_{i_{m-1} i_m} = \cdots \\
&\quad = p_{i_{n+1} i_n} p_{i_{n+2} i_{n+1}} \cdots p_{i_m i_{m-1}} \mu_{i_m} = \mu_{i_m} p_{i_m i_{m-1}} p_{i_{m-1} i_{m-2}} \cdots p_{i_{n+1} i_n} \\
&\quad = P(\omega: \xi_n(\omega) = i_m, \xi_{n+1}(\omega) = i_{m-1}, \cdots, \xi_m(\omega) = i_n).
\end{aligned}$$

于是我们有

**命题 3.21** $\xi$ 为以 $\underset{\sim}{\mu}$ 为可逆平稳分布的可逆马氏链当且仅当从初分布 $\underset{\sim}{\mu}$ 出发它成为平稳过程，而且对时间倒逆分布不变.

在 §3.3 例 3.6 中 $\beta_{ij} \equiv 1, \alpha_i = \frac{1}{n_i}$，其中 $n_i$ 表示从 $i$ 出发走一步可到达的车站总数.

又若网络能将全部状态联结成一体，则上述过程又一定只有一个互通常返类. 于是利用互通常返类的马氏链 $\xi_n$ 不变分布唯一性(§3.3 命题 3.13)，便得到

$$\pi_i = \mu_i = \alpha_i^{-1}\Big/\sum_i \alpha_i^{-1} = n_i\Big/\sum_i n_i.$$

这样 §3.3 例 3.6 的不变分布是

$$\pi_1 = \frac{2}{16} = \frac{1}{8},\quad \pi_2 = \frac{3}{16},\quad \pi_3 = \frac{2}{16} = \frac{1}{8},$$

$$\pi_4 = \frac{3}{16},\quad \pi_5 = \frac{2}{16} = \frac{1}{8},\quad \pi_6 = \frac{4}{16} = \frac{1}{4}.$$

这个结果正是例 3.6 中已经通过解方程得到的，但是这里我们利用了可逆性，从而不必去解方程，算法就大大简化了.

在物理上(3.34)称为**细致平衡条件**，这是因为由(3.34) 可得一个等价条件：对∀ $n$

$$\mu_i p_{ij}(n) = \mu_j p_{ji}(n), \tag{3.34$'$}$$

它的含义为：经过任何时间 $n$ ，对任意一对状态$(i,j)$，从 $i$ 到 $j$ 的概率 $\mu_i p_{ij}(n)$ 总与从 $j$ 到 $i$ 的概率 $\mu_j p_{ji}(n)$ 一样.

**例 3.7** (**$Z$ 上生灭链**) 设 $p_i + q_i = 1$，$p_i > 0, q_i > 0, (i = 0, \pm 1, \cdots)$，

$$p_{ij} = \begin{cases} p_i, & j = i + 1; \\ q_i, & j = i - 1; \\ 0, & \text{其他}. \end{cases}$$

这时只要令

$$\mu_i = \begin{cases} \prod\limits_{k=0}^{i-1}(p_k/q_{k+1}), & i > 0; \\ 1, & i = 0; \\ \prod\limits_{k=i+1}^{0}(q_k/p_{k-1}), & i < 0, \end{cases}$$

那么当 $\sum\limits_i \mu_i < \infty$ 时我们有

$$\tilde{\mu}_i p_{ij} = \tilde{\mu}_j p_{ji},$$

其中$\tilde{\mu}_i = \frac{\mu_i}{\sum_j \mu_j}$. 因而以 $P$ 为转移阵的马氏链是可逆的.

**例 3.8** 令

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix},$$

我们证明以 $P$ 为转移阵的马氏链 $\xi$ 不是可逆的. 事实上, 如果谬设 $\xi$ 是可逆的, 即(3.34)满足, 那么 $\underset{\sim}{\mu}$ 一定是 $P$ 的不变分布. 求解方程

$$\underset{\sim}{\mu} P = \underset{\sim}{\mu},$$

可知

$$\underset{\sim}{\mu} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right).$$

但是对 $i = 1$ 或 2, 我们有

$$\mu_i p_{i,i+1} - \mu_{i+1} p_{i+1,i} = \frac{1}{3} \cdot \frac{1}{2} - \frac{1}{3} \cdot 0 = \frac{1}{6} \neq 0,$$

即 $\underset{\sim}{\mu}$ 不满足(3.34). 这样就导致了矛盾. 从而 $\xi$ 不是可逆的.

可逆马氏链比起一般的正常返马氏链有两个显著的优点:

(1) 在可逆马氏链的状态集 $\mathscr{S}$ 中取一个子集 $\mathscr{S}_0$, 并将 $\mathscr{S}_0$ 外的状态都看成"反射壁", 则得到的新的马氏链仍为可逆的, 而且后者的可逆不变分布是前者的可逆不变分布在 $\mathscr{S}_0$ 上的限制. 但是一般不可约正常返马氏链作同样的"反射壁"新链后, 后者的不变分布并不是前者的不变分布在 $\mathscr{S}_0$ 上的限制.

(2) 可逆不变分布可以简单地用 $p_{ij}$ 明显表达, 但是一般的不变分布必须通过解方程 $\underset{\sim}{\mu} = \underset{\sim}{\mu} P$ 得到(当 $\mathscr{S}$ 的状态数 $N$ 很大时,

解这个方程的计算量是很惊人的).

以上两个优点分别对应下列两个命题:

**命题 3.22**(限制在状态子集 $\mathscr{S}_0$ 上的可逆不变分布) 设可逆马氏链 $\xi$ 以 $P$ 为转移阵,以 $\mu$ 为可逆不变分布. $\mathscr{S}_0$ 是状态集的子集. 令

$$\tilde{P}=(\tilde{p}_{ij})\ (i,j\in S_0),$$

其中 $i,j\in\mathscr{S}_0$,且

$$\tilde{p}_{ij}=\begin{cases}p_{ij}, & j\neq i;\\ p_{ii}+\sum\limits_{k\notin\mathscr{S}_0}p_{ik}, & j=i.\end{cases}$$

那么以 $\tilde{P}$ 为转移阵的马氏链 $\tilde{\xi}$ 为可逆的,它的可逆不变分布为

$$\tilde{\mu}=\{C\mu_i;i\in\mathscr{S}_0,\ C\equiv(\sum_{i\in\mathscr{S}_0}\mu_i)^{-1}\}.$$

**证明** 当 $j\neq i$ 时有 $\tilde{p}_{ij}=p_{ij}$,所以

$$\tilde{\mu}_i\tilde{p}_{ij}=C\mu_ip_{ij}=C\mu_jp_{ji}=\tilde{\mu}_j\tilde{p}_{ji}.$$

这说明 $\tilde{\mu}$ 是 $\tilde{\xi}$ 的可逆不变分布. 命题证毕.

另一方面,读者可自行验证,在例 3.2 中如果取 $\mathscr{S}_0=\{1,2\}$,那么 $P$ 的不变分布 $\mu=\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$,而 $\tilde{P}$ 的不变分布不是 $\tilde{\mu}=\left(\frac{1}{2},\frac{1}{2}\right)$,其原因在于 $\xi$ 不是可逆的.

**定理 3.23**(**Kolmogorov 可逆准则**) 互通的有限马氏链 $\xi$ 为可逆的当且仅当如下条件满足:

($K$) 对任意一个"状态环路" $i_1\to i_2\to\cdots\to i_m\to i_1$ 恒有

$$p_{i_1i_2}p_{i_2i_3}\cdots p_{i_{m-1}i_m}p_{i_mi_1}=p_{i_mi_{m-1}}p_{i_{m-1}i_{m-2}}\cdots p_{i_2i_1}p_{i_1i_m}.$$

这时对任意取定状态 $i_0\in\mathscr{S}_0$ 及状态 $i$,存在 $i_0$ 到 $i(\neq i_0)$ 的一条通路 $i_0\to i_1\to\cdots\to i_m\to i_{m+1}=i$,满足

$$v_i = \prod_{k=0}^{m} \left( \frac{p_{i_k i_{k+1}}}{p_{i_{k+1} i_k}} \right) > 0.$$

而且 $\xi$ 的可逆平稳分布 $\underset{\sim}{\mu} = (\mu_i)$ 为

$$\mu_i = \begin{cases} \alpha^{-1} v_i, & i \neq i_0; \\ \alpha^{-1}, & i \neq i_0, \end{cases}$$

其中 $\alpha = 1 + \sum_{i \neq i_0} v_i$ (可见 $\mu_i$ 可用 $p_{ij}$ 简单地表达).

注：一般互通马氏链也有对应的定理，请见 § 3.7 附录中的 3.

**证明** 必要性的证明：

首先，对于可逆平稳分布 $\underset{\sim}{\mu}$ 必存在 $i_0$ 使 $\mu_{i_0} > 0$ 而对其他的 $i \neq i_0$，由于互通性必存在 $n$ 使 $p_{i_0 i}(n) > 0$. 利用可逆条件(3.34)，我们有

$$\mu_{i_0} p_{i_0 i}(n) = \mu_i p_{i i_0}(n).$$

由左边恒正，故有 $\mu_i > 0$. 可见 $\mu_i > 0\ (\forall\ i)$.

任取一条通路 $i_0 \to i_1 \to \cdots \to i_m \to i$，利用(3.34)多次我们有

$$\mu_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_m i} = p_{i_1 i_0} \mu_{i_1} p_{i_1 i_2} \cdots p_{i_m i} = \cdots = p_{i_1 i_0} p_{i_2 i_1} \cdots p_{i i_m} \mu_i.$$

由左边恒正，推出右边恒正，并且对应的 $v_i$ 为 $\mu_i / \mu_{i_0} > 0$，

$$\alpha = 1 + \sum_{i \neq i_0} \frac{\mu_i}{\mu_{i_0}} = \frac{1}{\mu_{i_0}}.$$

充分性的证明.

首先我们注意当 $i_0$ 取定后，对 $i \neq i_0$ 在定理中定义的 $v_i$ 与 $i_0$ 到 $i$ 的通路取法无关.

事实上，设 $R_1, R_2$ 为 $i_0$ 到 $i$ 的两条不同通路：

$$R_1: i_0 \to i_1 \to \cdots \to i_{m-1} \equiv i,$$

$$R_2: i_0 \to i_1' \to \cdots \to i_{m+1}' \equiv i,$$

记 $R_1^-, R_2^-$ 为它们的反向通路，定义 $R_2^- \circ R_1$ 为如下的环状通路：$i_0$ 出发经 $R_1$ 到 $i$ 再经 $R_2^-$ 回到 $i_0$，类似地定义 $R_1^- \circ R_2$. 由(K)可知沿环路 $R_2^- \circ R_1$ 与沿环路 $R_1^- \circ R_2$ 的转移概率的乘积是相等的

(记沿 $R_1$ 及沿 $R_2$ 定义的 $v_i$ 分别为 $v_i^{(1)}, v_i^{(2)}$).

我们由此推出

$$v_i^{(1)}/v_i^{(2)} = 1.$$

即 $v_i^{(1)} = v_i^{(2)}$，从而可知 $v_i$ 的值不依赖通路的选取. 再由 $\mu_i$ 的定义推出 $\mu$ 是由 $P$ 唯一确定的. 下证 $\xi$ 可逆并以 $\mu$ 为可逆平稳分布. 由于互通性必存在一条从 $j$ 到 $i$ 的通路 $R$. 如果 $p_{ij} > 0$，记 $i \to j$ 的直接通路为 $R_1$. 那么转移概率沿环路 $R \circ R_1$ 的乘积为正. 由 (K) 得转移概率沿其反向回路 $R_1^{-1} \circ R^{-1}$ 的乘积也应为正. 这说明 $p_{ji} > 0$. 取 $i_0 = j, l = i$. 那么 $v_i = \frac{p_{ji}}{p_{ij}}, v_j = 1$. 从而 $v_i p_{ij} = v_j p_{ji}$. 由此 $\mu_i p_{ij} = \mu_j p_{ji}$. 又若 $p_{ij} = 0$，则 $p_{ji}$ 必须为 0(因为对 $p_{ji} > 0$ 用同样的推理会导致 $p_{ij} > 0$)，所以我们恒有 $\mu_i p_{ij} = \mu_j p_{ji}$，于是 $\xi$ 的可逆性得证.

**例 3.9** 设二维整格点上随机徘徊如下：

$$P_{(i,k),(j,l)} = \begin{cases} \alpha_{1,0}, & j = i+1, l = k; \\ \alpha_{0,1}, & j = i, l = k+1; \\ \alpha_{-1,0}, & j = i-1, l = k; \\ \alpha_{0,-1}, & j = i, l = k-1; \\ \alpha_{1,1}, & j = i+1, l = k+1; \\ \alpha_{-1,-1}, & j = i-1, l = k-1; \\ 0, & \text{其他}. \end{cases}$$

其中 $\sum_{n,m=-1}^{1} \alpha_{n,m} = 1$ $(\alpha_{0,0} = \alpha_{-1,1} = \alpha_{1,-1} = 0)$，$\alpha_{1,0} < \alpha_{-1,0}, \alpha_{0,1} < \alpha_{0,-1}$，$\alpha_{1,0}\alpha_{-1,-1}\alpha_{0,1} = \alpha_{0,-1}\alpha_{1,1}\alpha_{-1,0}$. 对于 $P$ 所对应的马氏链 $\xi$ 易验证(K)满足.

取 $i_0 = (0,0)$，定义 $v_{i_0} = 1$，$v_{(1,0)} = \frac{\alpha_{1,0}}{\alpha_{-1,0}}$，$v_{(0,1)} = \frac{\alpha_{0,1}}{\alpha_{0,-1}}$，$v_{(1,1)} = \frac{\alpha_{1,1}}{\alpha_{-1,-1}} = \frac{\alpha_{1,0}\alpha_{0,1}}{\alpha_{-1,0}\alpha_{0,-1}}$，归纳地令

$$\nu_{(n,m)}=\left(\frac{\alpha_{1,0}}{\alpha_{-1,0}}\right)^n\left(\frac{\alpha_{0,1}}{\alpha_{0,-1}}\right)^m,$$

那么

$$\sum_{n,m}\nu_{(n,m)}=\left(1-\frac{\alpha_{1,0}}{\alpha_{-1,0}}\right)^{-1}\left(1-\frac{\alpha_{0,1}}{\alpha_{0,-1}}\right)^{-1},$$

从而

$$\mu_{(n,m)}=\left(\frac{\alpha_{1,0}}{\alpha_{-1,0}}\right)^n\left(\frac{\alpha_{0,1}}{\alpha_{0,-1}}\right)^m\left(1-\frac{\alpha_{1,0}}{\alpha_{-1,0}}\right)\left(1-\frac{\alpha_{0,1}}{\alpha_{0,-1}}\right),$$

由此可知 $\xi$ 为可逆马氏链并以 $\mu$ 为可逆不变分布. 在这个例中求 $\mu$ 不必解方程,而直接得到表达式(事实上,去解方程 $\mu=\mu P$ 以求得 $\mu$ 的计算量要比上述方法的计算量大得多).

**3. Gibbs 样本生成法的进一步讨论**

(1) 多峰分布情形.

在 Gibbs 样本生成过程中,只有当马氏链 $\{\xi_n(\omega)\}$ 发展的次

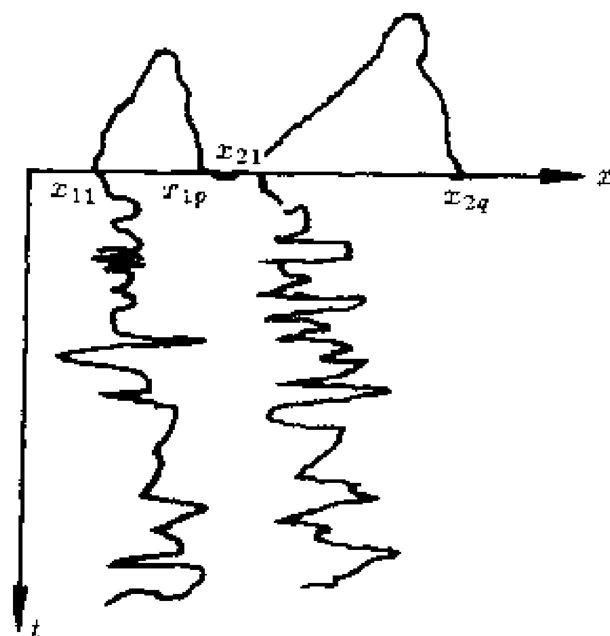



图 3.2

数充分大时，$\xi_n(\omega)$ 的分布才近似于 $\pi(x_1,x_2,\cdots,x_m)$．在实践中怎样才能判断 $n$ 是否已经足够大了呢？通常的判断标准似乎是看 $\xi_n(\omega)$ 的分布是否比较稳定了．可惜在事实上这种判断方法不尽正确．图 4.1 中给出了两峰的 $\pi(x)$ 例子．

从 $x_{11}$ 与 $x_{22}$分别出发的两条轨道分别稳定在 $E_1=\{x_{11},\cdots,x_{1l}\}$ 与 $E_2=\{x_{21},\cdots,x_{2l}\}$ 两个不同的区域．而事实上这时 $\xi_n(\omega)$ 还远未达到"遍历"，它们分别反映的分布都还不是 $\pi(x)$．原因在于平稳分布有两个峰，且在两峰中间的概率非常小，以致使得从 $E_1$ 中出发能到达 $E_2$ 和从 $E_2$ 中出发到达 $E_1$ 的概率都极为微小，从而在一条轨道上要从 $E_1$ 到 $E_2$ 就要化费十分长的时间．在实践上这个达到"遍历"的充分大的 $n$ 在现实生活中甚至可能不会在有意义的时间内达到．发现与使用 Gibbs 样本生成法的先驱往往希望通过选取多个初始值，得到多条轨道以解决这一问题．但是这也不能完全解决问题．例如在前面的例子中的两条轨道各自只局限在不同的区域 $E_1$ 与 $E_2$ 中，它们的分布根本反映不出 $\pi(E_1)$ 与 $\pi(E_2)$ 的比例．因而这样合在一起的样本从大局上一般不能期望从它得到 $\pi(x)$．

下面我们提出一种解决上述问题的方法——可逆马氏链 Gibbs 样本生成法，即通过可逆马氏链的发展来近似得到遵从分布 $\pi(x_1,x_2,\cdots,x_m)$ 的样本．如果 $(p_{xy})(p_{xy}=p(x,y))$ 是一个可逆马氏链的转移矩阵，那么

$$\pi(x)p(x,y)=\pi(y)p(y,x). \tag{3.35}$$

从可逆马氏链的特性就可以解决上面不能解决的各个峰区上 $\pi(\cdot)$ 的比例问题．下面我们仍以双峰时两条轨道的情形阐述确定比例分配的方法．

设分别在 $E_1$ 与 $E_2$ 上的两条轨道分别稳定在 $E_1$ 及 $E_2$ 上的样本分布为 $\hat{\pi}_1(x)(x\in E_1)$ 及 $\hat{\pi}_2(x)(x\in E_2)$．由可逆马氏链的性质，它们所分别代表的 $\{\xi_n(\omega):n\geqslant 0\}$ 限制在 $E_1$ 与 $E_2$ 上不变分布 $\pi_1(x)$ 与 $\pi_2(x)$ 应该分别与 $\{\pi(x):x\in E_1\}$ 及 $\{\pi(x):x\in E_2\}$ 成比

例，即

$$\frac{\hat{\pi}_i(x)}{\pi(x)} \approx \frac{\pi_i(x)}{\pi(x)} = \frac{1}{\pi(E_i)} = \text{常数} C_i \quad (\forall\, x \in E_i), \tag{3.36}$$

于是为了得到分布 $\pi(\cdot)$ 的样本，只要按 $\pi(E_i)=C_i^{-1}(i=1,2)$ 的比例从上述所得的 $E_1,E_2$ 中的样本抽取．也即只须知道 $C_i$．

确定 $C_i$ 的方法如下：

对 $x_1,x_2$ 分别 $\in E_1$ 与 $E_2$，

$$\frac{C_1}{C_2} = \frac{\pi_1(x_1)\pi(x_2)}{\pi_2(x_2)\pi(x_1)} = \frac{\pi_1(x_1)}{\pi_2(x_2)}\left(\frac{p(x_1,x_2)}{p(x_2,x_1)}\right),$$

其中 $\pi_i(x_i)(i=1,2)$ 应满足：对 $\forall\, x,y \in E_i$，

$$1 = \sum_{x\in E_i}\pi_i(x) = \sum_{x\in E_i}\frac{\pi_i(y)p(y,x)}{p(x,y)} = \pi_i(y)E_y\left\{\frac{1}{p(\xi_i,y)}\right\},$$

故
$$\pi_i(y) = \left[E_y\left\{\frac{1}{p(\zeta,y)}\right\}\right]^{-1}.$$

而 $E_y\left\{\frac{1}{p(\xi_1,y)}\right\} \doteq \frac{1}{N}\sum_{k=1}^{N}\frac{1}{p(\xi_1^k,y)}$（其中 $\{\xi_1^k\}$ 是 $\xi_1$ 的 $N$ 个独立模拟样本）．

上式的近似只需作 $N$ 次一步马氏链的模拟．它比马氏链的分布近似于不变分布要求的步数要少得很多．

余下的问题是从 $\pi(x)$ 构造马氏链，使它们以 $\pi(x)$ 为可逆不变分布．

由第二段的讨论我们可以构造马氏链 $(p(x,y))$ 以 $\pi(x)$ 为不变分布．再令

$$\hat{p}(x,y) = \frac{1}{2}\left\{p(x,y) + \frac{\pi(y)p(y,x)}{\pi(x)}\right\},$$

容易检查 $(\hat{p}(x,y))$ 也是转移阵，而且满足

$$\pi(x)\hat{p}(x,y) = \pi(y)\hat{p}(y,x),$$

即它以 $\pi$ 为可逆不变分布．可见以 $(\hat{p}(x,y))$ 为转移阵的马氏链就是我们所要的．

为了便于操作，我们还希望 $\hat{p}(x,y)$ 具有与(3.33)类似的形

式，即每次只改一个分量的 $m$ 个“转移概率”的乘积. 为此我们注意：

$$
\begin{aligned}
\frac{\pi(y)p(y,x)}{\pi(x)} &= \frac{\pi(y_1,\cdots,y_m)\pi(x_1\mid y_2,\cdots,y_m)\cdots\pi(x_m\mid x_1,\cdots,x_{m-1})}{\pi(x_1,\cdots,x_m)} \\
&= \frac{\pi(y_1,\cdots,y_m)}{\pi(x_1,\cdots,x_m)}\frac{\pi(x_1,y_2,\cdots,y_m)}{\int\pi(y,y_2,\cdots,y_m)\mathrm{d}y}\cdots\frac{\pi(x_1,\cdots,x_{m-1},y_m)}{\int\pi(x_1,\cdots,x_{m-2},y,y_m)\mathrm{d}y} \\
&\quad\times\frac{\pi(x_1,\cdots,x_m)}{\int\pi(x_1,\cdots,x_{m-1},y)\mathrm{d}y} \\
&= \pi(y_1\mid y_2,\cdots,y_m)\cdots\pi(y_{m-1}\mid x_1,\cdots,x_{m-2},y_m)\pi(y_m\mid x_1,\cdots,x_{m-1}),
\end{aligned}
\tag{3.33'}
$$

即 $\frac{\pi(y)p(y,x)}{\pi(x)}$ 具有与(3.33)类似的形式，与(3.33)不同处只在于所有条件分布中的变量次序倒了一下. 而 $\hat{p}(x,y)$ 的意思是分别以 $\frac{1}{2}$ 的概率取(3.33)形式与(3.33)$'$的形式. 所以从模拟程序的实现来看这并未增加多少计算量. 其实只增加了一次二进制随机选择.

(2) 一元(条件)分布的简单化样本计算法.

以下不妨假定该分布具有密度(在离散情形可以完全类似地仿作).

前面所述的 Gibbs 采样方法最终归结为把多元样本问题化为生成一元条件分布的样本问题. 为了得到以 $F(x)$ (一元分布函数)为分布的样本函数，只要生成一组 $[0,1]$ 上 i.i.d. 的均匀分布的随机变量 $\{Z_1,Z_2,\cdots,Z_n\}$. 当 $F(X)$ 严格单调时，$\{F^{-1}(Z_1),\cdots,F^{-1}(Z_n)\}$ 就是分布为 $F(x)$ 为 i.i.d. 的随机变量. 由于在实际计算中通常都是知道 $F(x)$ 的分布密度(此密度一般连续)，要得到 $F^{-1}$ 就需要一定的计算量，特别由于在 Gibbs 样本方法中这种步骤常需要重复成千上万次，所以简化一元分布的实际计算方法是极有价值的.

下面介绍一种简单易行的生成以 $f(x)$ 为分布密度的样本的方法，称为 Von Neumann 的**拒收原则**(**rejection principle**，也称 **importance sampling**). 其方法为：

① 把分布密度 $f(x)$ "改为"与它有相同取值范围的简单分布密度(如指数分布，正态分布，均匀分布或它们的混合分布) $\rho(x)$，只要满足：$\exists\, C$ 使

$$f(x) \leqslant C\rho(x) (\triangleq g(x)).$$

② 独立地生成一组分布为 $\rho(x)$ 为样本 $\{Z_1, \cdots, Z_n\}$ 及 $[0,1]$ 的均匀样本 $\{U_1, \cdots, U_n\}$.

③ 如果

$$U_i > f(Z_i)/g(Z_i),$$

则剔除 $Z_i$，余下的就认为是一组以 $f(x)$ 为分布密度的样本.

下面我们证明留下的 $Z_i$ 确实以 $f(x)$ 为分布密度.

注意到经过剔除后的 $Z_i$ 的分布实际上是如下的条件分布

$$\begin{aligned}
&P(Z \leqslant x \mid f(Z)/g(Z) \geqslant U) \\
&= \frac{P(Z \leqslant x, f(Z)/g(Z) \geqslant U)}{P(f(Z)/g(Z) \geqslant U)} \\
&= \frac{\displaystyle\int_{-\infty}^{x} P\left\{U \leqslant \frac{f(z)}{g(z)}\right\} \rho(z)\mathrm{d}z}{\displaystyle\int_{-\infty}^{\infty} P\left\{U \leqslant \frac{f(z)}{g(z)}\right\} \rho(z)\mathrm{d}z} \\
&= \int_{-\infty}^{x} \frac{f(z)}{g(z)}\rho(z)\mathrm{d}z \Big/ \int_{-\infty}^{\infty} \frac{f(z)}{g(z)}\rho(z)\mathrm{d}z.
\end{aligned}$$

利用 $\dfrac{\rho(z)}{g(z)} = \dfrac{1}{C}$ 及 $\displaystyle\int_{-\infty}^{\infty} f(z)\mathrm{d}z = 1$，便得

$$P(Z \leqslant x \mid U \leqslant f(z)/g(z)) = \int_{-\infty}^{x} f(z)\mathrm{d}z = F(x).$$

而 $[0,1]$ 均匀随机变量 $U_i$ 可以直接用伪随机数生成. 指数分布的随机样本，正态分布的随机样本都是易于得到的. 请读者自己思考对于正态分布如何简单地生成样本.

(3) 简单应用.

**例** $(X_1,X_2)\sim N\left(0,\begin{pmatrix}\sigma_1^2 & r\sigma_1\sigma_2\\ r\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$，近似地求数 $a$，使

$$P\left(\frac{\sin^2 X_1+\sin^2 X_2}{1+\mathrm{tg}^2 X_1 \mathrm{tg}^2 X_2}\geqslant a\right)\approx\epsilon.$$

记括号内的随机变量为 $\psi(X_1,X_2)$，由 Gibbs 样本生成法可以生成一组 $(X_1,X_2)$ 的样本 $(X_1^{(k)},X_2^{(k)})$，$k=1,2,\cdots,n$. 于是 $\psi(X_1,X_2)$ 的样本分布为

$$\#\{k:\psi(X_1^{(k)},X_2^{(k)})\leqslant x\}/n.$$

这样我们可以取

$$a\approx\max\{x:\#\{k:\psi(X_1^{(k)},X_2^{(k)})\geqslant x\}\geqslant n\epsilon\}.$$

这里的思想是直接用样本(实际上是附加样本)近似地确定临界值.

注：运用拒受法剔除保留一些样本，有时可能因剔除太多而在计算上不经济，人们就考虑各种改进，在应用领域中出现的各种 Monte Carlo 采样法，Las Vegas 采样法等就应运而生，而且在不同的实用领域各有所长.

## §3.7 附　　录

1. 本章中的定义与定理除特别声明(如定理 3.8，定理 3.13)外，一般也都适用于可数状态的马氏链. 我们只对有限状态情形作了证明是为了回避测度论等较为高深的数学工具，而突出该问题与定理的主要思路与应用. 对数学的严格性有兴趣的读者可在《随机过程论》等类似水平的书中找到相应定理的严格叙证.

2. 引理 3.14 的证明.

记

$$A_n=\{1,2,\cdots,n\},$$

于是

$$\sum_{k=1}^{n}a_k b_{n,k}=\sum_{k=1}^{\infty}a_k b_{n,k}\mathbf{1}_{A_n}(k).$$

定义测度 $\mu$ 满足

$$\mu(\{k\}) = a_k,\ \mu(R \backslash N) = 0,\ N = (\{1,2,\cdots\}).$$

它是有限测度$\left(因为\sum_k a_k < \infty\right)$. 再令

$$b_n(x) = b_{n,k}\mathbf{1}_{A_n}(k)\ (k \leqslant x < k+1)\quad (k \geqslant 0).$$

由于 $b_{n,k} \leqslant C$，所以 $b_n(x)$ 一致有界. 这样，上面的无穷级数

$$\sum_{k=1}^{\infty} a_k b_{n,k}\mathbf{1}_{A_n}(k) = \int_0^{\infty} b_n(x)\mu(\mathrm{d}x).$$

利用有界收敛定理，当 $n \to \infty$ 时上式右方趋于

$$\int_0^{\infty} \lim_{n\to\infty} b_n(x)\mu(\mathrm{d}x) = \sum_{k=1}^{\infty} a_k \lim_{n\to\infty} b_{n,k}\mathbf{1}_{A_n}(k) = \sum_{k=1}^{\infty} a_k b_k.$$

请者也可以利用数学分析方法直接证明引理 3.14.

3. 对称马氏链. 转移阵 $\boldsymbol{P} = (P_{ij})$ 对应的时齐可数马氏链 $\xi$ 称为对称的，如果存在一个正数列 $\{\mu_i\}(i \in \boldsymbol{S})$ 满足

$$\mu_i P_{ij} = \mu_j P_{ji}\quad (\forall\ i,j \in \mathscr{S}).$$

这时 $\{\mu_i\}$ 称为对称化(或配称)数列$\left(这时也可能有\sum_i \mu_i = +\infty\right)$.

如果对称化数列 $\{\mu_i\}$ 还满足

$$\sum_i \mu_i < \infty,$$

那么对于

$$\tilde{\mu}_i \triangleq \mu_i \Big/ \sum_i \mu_i,$$

我们仍有

$$\tilde{\mu}_i P_{ij} = \tilde{\mu}_j P_{ji}.$$

可见这时 $\{\tilde{\mu}_i\}$ 是 $\boldsymbol{P}$ 的可逆平稳分布，因而马氏链 $\xi$ 是可逆的. 于是我们看到，对称马氏链是比可逆马氏链更为广的概念. 一个不可约对称马氏链为可逆马氏链当且仅当其对称化数列 $\{\mu_i\}$ 满足 $\sum_i \mu_i < \infty$. 而对于有限状态的马氏链，可逆性与对称性是一样的.

一个以$\{\mu_i\}$为对称化数列的对称马氏链的转移阵$\boldsymbol{P}$有以下性质：令

$$(\pi_{ij}) \triangleq \begin{pmatrix} \sqrt{\mu_1} & & 0 \\ & \sqrt{\mu_2} & \\ 0 & & \ddots \end{pmatrix} \boldsymbol{P} \begin{pmatrix} \frac{1}{\sqrt{\mu_1}} & & 0 \\ & \frac{1}{\sqrt{\mu_2}} & \\ 0 & & \ddots \end{pmatrix},$$

即

$$\pi_{ij} \triangleq \sqrt{\mu_i} P_{ij} \frac{1}{\sqrt{\mu_j}},$$

则$(\pi_{ij})$是对称阵. 事实上，利用对称马氏链定义：

$$\begin{aligned}\pi_{ij} &= \sqrt{\mu_i} P_{ij} \frac{1}{\sqrt{\mu_j}} = \frac{1}{\sqrt{\mu_i}} \mu_i P_{ij} \frac{1}{\sqrt{\mu_j}} \\ &= \frac{1}{\sqrt{\mu_i}} \mu_j P_{ji} \frac{1}{\sqrt{\mu_j}} = \sqrt{\mu_j} P_{ji} \frac{1}{\sqrt{\mu_i}} = \pi_{ji}.\end{aligned}$$

下面我们进一步讨论有限状态的对称马氏链(此时一定是可逆马氏链).

设状态数为$N$. 那么对称矩阵$\pi=(\pi_{ij})$可以写成线性代数的标准形，即存在正交矩阵$\mathbf{U}$，使

$$\boldsymbol{\pi} = \boldsymbol{U}^{-1} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix} \boldsymbol{U} \quad (\lambda_i \text{ 为实数}, i=1,2,\cdots,N).$$

于是

$$\boldsymbol{P} \triangleq \begin{pmatrix} \frac{1}{\sqrt{\mu_1}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\mu_N}} \end{pmatrix} \boldsymbol{\pi} \begin{pmatrix} \sqrt{\mu_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\mu_N} \end{pmatrix}$$

$$= W^{-1}\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_N\end{pmatrix}W,$$

其中

$$W = U\begin{pmatrix}\sqrt{\mu_1} & & 0\\ & \ddots & \\ 0 & & \sqrt{\mu_N}\end{pmatrix}.$$

可见 $\lambda_1,\lambda_2,\cdots,\lambda_N$ 是 $\boldsymbol{P}$ 的特征值.

又由于 $\boldsymbol{\pi}$ 的对称性,$\boldsymbol{\pi}$ 具有 $N$ 个线性无关的特征向量 $\vec{\mu}_1,\cdots,\vec{\mu}_N$:

$$\boldsymbol{\pi}\vec{\mu}_i = \lambda_i\vec{\mu}_i,$$

可见对于

$$\vec{v}_i \equiv \begin{pmatrix}\frac{1}{\sqrt{\mu_1}} & & 0\\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\mu_N}}\end{pmatrix}\vec{\mu}_i,$$

有

$$\boldsymbol{P}\vec{v}_i = \lambda_i\vec{v}_i,$$

也就是 $\boldsymbol{P}$ 有 $N$ 个线性无关的特征向量. 现在设 $\vec{v}$ 是其中某一个 $\vec{v}_i$,它对应的 $\lambda_i$ 记为 $\lambda$,假定

$$\vec{v} = \begin{pmatrix}v^{(1)}\\ \cdots\\ v^{(N)}\end{pmatrix},\ \text{且}\ |v^{(i_0)}| = \max_j|v^{(j)}|.$$

我们不妨假定 $v^{(i_0)} = 1$ $\left(\text{否则用}\frac{\vec{v}}{v^{(i_0)}}\text{代替}\vec{v}\right)$. 于是

$$|\lambda| = |\lambda v^{(i_0)}| = \left|\sum_{j=1}^{N} P_{i_0 j} v^{(j)}\right| \leqslant \sum_{j=1}^{N} P_{i_0 j} = 1 .$$

此外我们显见 $\lambda=1$ 是 $\boldsymbol{P}$ 的特征值，它具有特征向量 $\mathbf{1}$. 这是因为

$$\boldsymbol{P}\mathbf{1} = \mathbf{1}.$$

综上所述，我们有：

对称马氏链的转移矩阵 $\boldsymbol{P}$ 的特征值都是实的，其绝对值不超过 1，而且 1 一定是 $\boldsymbol{P}$ 的特征值. 此外 $\boldsymbol{P}$ 具有 $N$ 个线性无关的特征向量. 由此还可知 $\boldsymbol{P}-\boldsymbol{I}$ 的特征值都是非正的.

而一般有限马氏链的转移矩阵 $\boldsymbol{P}$ 的特征值却未必全是实的. 但是类似地可证它们的模仍不超过 1，而且 1 仍然是特征值. 特别，如果 $\boldsymbol{P}$ 只有一个互通类，则有著名的 Frobeniue-Perron 定理：特征值 1 是单重的.

用 $\boldsymbol{P}$ 来判定有限马氏链是否可逆的 Kolmogorov 条件也是可数状态的对称马氏链的转移阵的充要条件. 事实上，Kolmogorov 原来的充要条件就是刻画可数状态的马氏链的对称性的.

最后，我们举一个对称而又不是可逆的马氏链的例子，它就是简单对称随机徘徊. 它的对称化数列是 1，即 $\mu_i \equiv 1$，显见 $\sum_i \mu_i = \infty$，所以它不是可逆平稳分布. 从这个例子可看到仅仅研究可逆马氏链是不够的.

4. 设 $\{\mu_i\}$ 是有限状态对称马氏链的转移阵 $\boldsymbol{P}$ 的对称化数列. 记

$$f(x) \triangleq \sum_{i,j=1}^{N} \mu_i x_i (p_{ij} - \delta_{ij}) x_j = \sum_{i,j=1}^{N} x_i (\mu_i p_{ij} - \mu_i \delta_{ij}) x_j.$$

由于 $\mu_i p_{ij} = \mu_j p_{ji}$，所以 $f(x)$ 是对称二次型，称为该马氏链的能量二次型，也称为马氏链的 Dirichlet 型. Dirichlet 型的概念可推广到很一般的状态空间（可数状态空间，$R^d$，甚至无穷维的空间）以及时间参数为连续的情形. 由于在那里的马氏过程可能极为复杂，无法象本章中那样直接去研究，Dirichlet 型就成为研究马氏过程的一个重要工具. 近来 Dirichlet 型与其相关的马氏过程已成为马

氏过程理论中的一个极为重要的分支.

## 习　　题

1. 设$\{\xi_n: n\geqslant 0\}$为$Z^1$上对称的简单随机徘徊,证明$\{|\xi_n|: n\geqslant 0\}$为时齐的马氏链,并求它的转移阵.

2. 设$\{\xi_n: n\geqslant 0\}$为马氏链,$\eta_n=(\xi_n,\xi_{n-1}-\xi_n)$,求证:

a) $\{\eta_n\}$为马氏链;

b) $\{\zeta_n\}$为马氏链,其中$\zeta_n=(\xi_n,\xi_{n+1},\cdots\xi_{n+p})$.

3. 设$\{\eta_n: n=0,\pm 1,\pm 2,\cdots\}$为独立同分布列,且

$$\eta_n \sim \begin{bmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

又设$\{\zeta_n, n\geqslant 0\}$为从0出发的$Z^1$上的对称的简单随机徘徊,且$\{\zeta_n\}$与$\{\eta_n\}$独立. 令

$$\xi_n = \eta_{\zeta_n},$$

称$\{\xi_n: n\geqslant 1\}$为在随机徘徊中观察到的随机场景.

a) 求$E\xi_n$;

b) 证明

$$E(\xi_n\xi_{n+m}) = \begin{cases} = 0, & m \text{ 为奇数} \\ \approx \sqrt{\dfrac{2\pi}{m}}, & m \text{ 为偶数,且 } m\to\infty. \end{cases}$$

4. 用转移阵给出一个取整值的时齐马氏链为简单随机徘徊的充要条件.

5. 设$\epsilon_n$为独立同分布列,$\epsilon_n \sim \begin{bmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$. 设某商品在整时间供应、出售并调价,其在时刻$n$的价格$\xi_n$在1元、2元至5元(整数)按下列方式浮动(调价政策):

$$\xi_{n+1}-\xi_n=\begin{cases}1, & 若 D_n-C_n>0;\\ 0, & 若 D_n-C_n=0;\\ -1, & 若 D_n-C_n<0.\end{cases}$$

其中 $C_n$、$D_n$ 分别为时刻 $n$ 时的供应量与需求量，满足

$$\begin{cases}D_n=10-\xi_n,\\ C_n=7+\xi_n.\end{cases}$$

a）求证 $\xi_n$ 是马氏链；

b）计算一步转移阵和两步转移阵；

c）计算不变分布 $\pi$ 和平衡价格 $E_\pi\xi_n$.

6. 设水库的储量都是整值的，其最大容量为 $N$（整数）吨，每日的进水量是独立同分布的整值随机变量（记第 $n$ 日的进水量为 $\varepsilon_n$ 吨）. 又设 $P(\varepsilon_n=k)=a_k$. 平时当水库不空时，每日从坝释放水 1 吨（空时不释放），但当水库水超过最大容量 $N$ 时，就释放超过最大容量的部分. 求证在第 $n$ 天水库中水量 $\xi_n$ 是一个马氏链，并求其转移阵及不变分布.

7. 试证明有限状态马氏链不可能有零常返状态.

8. 设 $\xi_n=(\xi_n^{(1)},\xi_n^{(2)})$ 是从原点出发的对称二维简单随机徘徊. 记

$$\tau=\inf\{n:|\xi_n^{(1)}|+|\xi_n^{(2)}|=2\},$$

请直观地说出 $\xi_{\tau+n}$ 的分布（$n\geqslant 0$）.

9. 盒中有号码为 1 至 $N$ 的票各一张，每次任取一张，记下号码后放回，记

$$T_j=\inf\{n:前\ n\ 次抽到的票中恰有\ j\ 种不同号码\}\ (T_1=1),$$

$$\tau_j=\inf\{n:在第\ n\ 次抽取时恰好抽到\ j\ 号票\}.$$

a）求 $T_j$，$\tau_j$ 的分布；

b）求证 $T_1,T_2-T_1,\cdots,T_N-T_{N-1}$ 独立；

c）求证：$\exists\ a>0$，使得 $P(T_N>m)\leqslant Ne^{-am/N}$；

d）求证 $P(T_N>m)=\sum\limits_{k=1}^{N}(-1)^{k-1}C_N^k\left(1-\dfrac{k}{N}\right)^m$.

10. 设 $x_n$ 是非负的独立同分布随机变量列，又设其分布函数连续，定义记录$\{x_n\}$列中最大值出现的随机时刻：$R_1=1, R_{k+1}=\inf\{n>R_k: x_n\geqslant\max(x_1,\cdots,x_{n-1})\}$. 求证$\{R_k\}$是马氏链，求出它的转移阵. 又问$\{T_k\}$及$\{(R_k,T_k)\}$是否马氏链($T_k\triangleq R_{k+1}-R_k$)?

11. 带号码 1 至 $N$ 的 $N$ 个球被放在 $A$，$B$ 两个盒中，每次等概地随机选一个号码，并与选号码独立地以概率 $p$ 选盒 $A$、以概率 $q=1-p$ 选盒 $B$，如果选到的号码对应的球不在被选到的盒内，则把此球移到被选中的盒中，否则保持原状态不变. 记在第 $n$ 次选取后 $A$ 中的球数为 $\xi_n$，求证$\{\xi_n\}$是马氏链，并求出它的转移阵及不变分布.

12. $N$ 个位置均匀放在圆周上，一个质点每隔单位时间以概率 $p$ 逆钟方向移动一个位置，以概率 $q=1-p$ 顺钟方向移动一个位置. 求此马氏链的转移阵及不变分布.

13. Doeblin 独立耦合方法.

设 $\boldsymbol{P}$ 为有限状态空间$\{1,2,\cdots,N\}$上的转移矩阵，即 $\boldsymbol{P}=(p_{ij}), p_{ij}\geqslant 0, \sum\limits_j p_{ij}=1, 1\leqslant i\leqslant N$，如果它满足：$\exists\ m$，使 $\boldsymbol{P}^m\equiv(p_{ij}^{(m)})$有 $p_{ij}^{(m)}>0$. 设 $\tilde{\pi}$ 为 $\boldsymbol{P}$ 的不变分布：$\tilde{\pi}\boldsymbol{P}=\tilde{\pi}$($\tilde{\pi}$ 看成行向量)，又$\{\xi_n\}$，$\{\eta_n\}$是相互独立且都以 $\boldsymbol{P}$ 为转移阵的马氏链，满足 $\xi_0\sim\tilde{\pi}$，$\eta_0\equiv i$. 记随机时间 $T=\inf\{n:\xi_n=\eta_n\}$.

a) 证明：

$$P(\xi_{km}\neq\eta_{km}\mid\xi_{(k-1)m}\neq\eta_{(k-1)m},\cdots,\xi_m\neq\eta_m)$$
$$\leqslant(1-N\delta^2)(\delta\triangleq\min_{i,j}p_{ij}^{(m)}),$$
$$P(T>km)\leqslant(1-N\delta^2)^k,$$
$$P(T<\infty)=1.$$

b) 证明：$(\xi_n,\eta_n)$是马氏链，$T$ 是它的停时.

c) 证明如下定义的$\{\zeta_n\}$是马氏链，它以 $\boldsymbol{P}$ 为转移阵，以 $\tilde{\pi}$ 为不变分布：

$$\zeta_n = \begin{cases} \eta_n & n \leqslant T; \\ \xi_n, & n > T. \end{cases}$$

d) 证明 $|P(\xi_n = j) - P(\zeta_n = j)| \leqslant P(T \geqslant n)$.

e) 证明(由(1)):

$$|p_{ij}^{(n)} - \pi_j| \leqslant (1 - N\delta^2)^{\frac{n}{m}-1}.$$

这一习题用独立耦合方法给出了有限马氏链的渐近于不变分布的性质的证明. 耦合方法、尤其是非独立的耦合方法, 如今已成为随机过程与统计力学中的重要方法.

14. 双随机转移阵.

有限状态的转移阵 $\boldsymbol{P}$ 称为双随机的, 如果它的转置 $\boldsymbol{P}^t$ 也是转移阵. 对于只有一个不可约互通类的双随机阵, 请求出它的不变分布.

15. 设 $\boldsymbol{P}$ 为有限的双随机阵, 则离散的均匀分布是 $\boldsymbol{P}$ 的配称分布当且仅当 $\boldsymbol{P}$ 为对称.

16. 对下列 $\boldsymbol{P}$ 指出哪些状态是常返的和暂态的:

a) $p_{i,i+1}=\dfrac{1}{i+1}, p_{i,0}=\dfrac{i}{i+1}\quad (i=0,1,2,\cdots)$;

b) $p_{0,1}=1, p_{i,0}=\dfrac{1}{i+1}, p_{i,i-1}=\dfrac{i}{i+1}\quad (i=0,1,2,\cdots)$;

c) $p_{i,0}=p\quad (0<p<1), p_{i,i+1}=q\equiv 1-p\quad (i=0,1,2,\cdots)$.

17. 某机器可处于优良、一般、待修三种状态(分别记成 1, 2, 3). 该机器在优良状态下停留 $\lambda_1$ 时间后就分别以 $\dfrac{3}{4}, \dfrac{1}{4}$ 概率转入一般状态和待修状态; 在一般状态工作 $\lambda_2$ 时间后就转入待修状态; 在待修状态修理 $\lambda_3$ 时间后就分别以概率 $\dfrac{2}{3}$ 与 $\dfrac{1}{3}$ 转入优良状态和一般状态. 求机器分别处在三种状态中的时间之比.

18. 设 $\{\xi_n\}$ 是不可约的非常返(暂态)马氏链, 记

$$G_{ij} = \sum_{n=0}^{\infty} p_{ij}^{(n)},$$

求证:

a) $0 < G_{ij} < \infty$;

b) $G_{ij} \leqslant G_{jj}$;

c) $\dfrac{G_{ij}}{G_{ii}} \leqslant \dfrac{G_{kj}}{G_{ki}}$;

d) $\sup\limits_{k} \dfrac{G_{kj}}{G_{ki}} \leqslant \dfrac{G_{jj}}{G_{ji}}$;

d) $\sup\limits_{k} \dfrac{G_{jk}}{G_{ik}} < \infty$.

19. 蜘蛛与蝇分别处于 1,2 两个位置之一,如相遇则捕食. 设蜘蛛与蝇分别按下列转移规则独立行动:

$$\begin{pmatrix} p & q \\ q & p \end{pmatrix}, \quad \begin{pmatrix} p' & q' \\ q' & p' \end{pmatrix} (0 < p, p' < 1, \quad p + q = p' + q' = 1).$$

试用一个三状态马氏链表示捕食情况,并求出:

a) $n$ 步后恰与初位置一样的概率;

b) 平均捕食时间.

20. 设

$$d_{ij} = d_{ji} > 0 \quad (i, j = 1, 2, \cdots, n),$$

$$p_{ij} = \begin{cases} 0, & i = j; \\ d_{ij} \Big/ \sum\limits_{k \neq j} d_{ik}, & i \neq j. \end{cases}$$

求证 $\boldsymbol{P}$ 是可逆马氏链的转移阵,并求$\lim\limits_{n \to \infty} \boldsymbol{P}^n$.

21. 设 $\boldsymbol{P}$ 是可逆马氏链的转移阵,其可逆不变分布为 $\tilde{\pi}$ ,记

$$\bar{p}_{ij} = \begin{cases} p_{ij} \Big/ \sum\limits_{j=1}^{n} p_{ij}, & 1 \leqslant i, j \leqslant n; \\ 0, & \text{其他}. \end{cases}$$

证明 $\bar{\boldsymbol{P}}$ 是可逆马氏链的转移阵,并证明它的可逆不变分布 $\pi$ 为

$$\pi_i = \frac{\tilde{\pi}_i \sum\limits_{j=1}^{n} p_{ij}}{\sum\limits_{i,j=1}^{n} p_{ij} \tilde{\pi}_i}.$$

22. 证明二维对称简单随机徘徊是常返的，而三维以上对称简单随机徘徊是非常返(暂态)的 .

23. Dobrushine 收缩不等式.

记 $\alpha=(\alpha_1,\alpha_2,\cdots)$，$\|\alpha\|=\sum_{i=1}^{\infty}|\alpha_i|$. 设 $\boldsymbol{P}$ 为马氏链的转移阵，$\mu,\nu$ 均为初分布，求证：

$$\|\mu\boldsymbol{P}-\nu\boldsymbol{P}\|\leqslant\frac{1}{2}C(\boldsymbol{P})\|\mu-\nu\|,$$

其中

$$C(\boldsymbol{P})\triangleq\sup_{i\neq k}\sum_j|p_{ij}-p_{kj}|.$$

提示：令 $\alpha^{\pm}\triangleq(\mu-\nu)^{\pm}$ $(\alpha^{\pm}\triangleq(\alpha_1^{\pm},\alpha_2^{\pm},\cdots))$.

(1) 证明 $\|\alpha^{\pm}\|=\|\alpha\|/2$;

(2) 证明

$$\sum_i\alpha_i p_{ij}=\sum_{i,k}\frac{\alpha_i^{+}\alpha_k^{-}}{\frac{1}{2}\|\alpha\|}p_{ij}-\sum_{i,k}\frac{\alpha_i^{-}\alpha_k^{+}}{\frac{1}{2}\|\alpha\|}p_{kj}=\sum_{i\neq k}\frac{\alpha_i^{+}\alpha_k^{-}}{\frac{1}{2}\|\alpha\|}(p_{ij}-p_{kj}).$$

24. 利用 Dobrushine 收缩不等式证明：若有限状态转移阵 $\boldsymbol{P}$ 满足 $\min_{i,j}p_{ij}>0$，则对任意初分布 $\mu$，$\lim_{n\to\infty}\mu\boldsymbol{P}^n$ 存在而且与 $\mu$ 无关，同时有

$$\|\mu\boldsymbol{P}^n-\pi\|\leqslant Ce^{-\alpha n},$$

其中 $\pi$ 为极限分布，$C,\alpha>0$.

25. 证明 $C(\boldsymbol{P}^{n+m})\leqslant\frac{1}{2}C(\boldsymbol{P}^n)C(\boldsymbol{P}^m)$，其中

$$C(\boldsymbol{P})\triangleq\sup_{i\neq k}\sum_j|p_{ij}-p_{kj}|.$$

26. 利用一个[0,1]均匀随机变量 $u$ 生成混合分布随机数.

又若分布函数 $F_i(x)$ $(i\leqslant N)$ 严格递增，$p_i>0$ $(i\leqslant N)$ 且 $\sum_{i=1}^{N}p_i=1$，又 $t_0=0<t_1<\cdots<t_N=1$，求证

$$\sum_{i=1}^{N} F_i^{-1}\left(\frac{u-t_i}{p_i}\right) I_{(t_{i-1},t_i]}(u) \sim \sum_{1}^{N} p_i F_i.$$

请用此方法构造遵从混合指数分布随机数.

# 第四章　马氏链的应用与特例

马氏链的应用非常广泛，它涉及物理、生物、化学、神经网络、信息、经济、金融等许多领域. 本章中，我们将在第二章关于马氏链的基本概念与知识的基础上，列举出马氏链的各种应用模式，以期给读者将马氏链应用到自己相关领域作一点启示.

由于实际问题往往是十分复杂的，如不对其进行适当的简化，不仅可能使问题因过于复杂而无法解决，而且也会使主要矛盾被纷繁的枝节掩盖而难以理解. 因而，在本章中，我们对许多实际问题进行必要的抽象和简化，以便于读者理解其实质. 于是，也就不可避免地会使问题可能与实际有一定的差距. 在读者应用马氏链去解决实际问题时，还需具体问题具体分析，根据实际情况去变通，灵活地应用各种模型.

## §4.1　Galton-Watson(GW)简单分支过程

### 1. 概念与模型

本节的数学模型来源于中子、细胞或微生物等(统一简称为粒子)的分裂与死亡的随机现象. 这里每一代裂变后的粒子的总数决定下一代粒子的总数的分布.

设$\{\xi_{n,k}: n,k\geqslant 1\}$是一族相互独立同分布取非负整数值的随机变量. 令

$$f_i = P(\xi_{n,k} = i)\ (i = 0,1,\cdots), \tag{4.1}$$

$\xi_{n,k}$表示第$n$代的第$k$个粒子分裂成的粒子数，$\xi_{n,k}=0$意味着此粒子死亡. 于是第$n+1$代的粒子总数是

$$X_{n+1} = \sum_{k=1}^{X_n} \xi_{n,k}. \tag{4.2}$$

设开始是一个粒子,即 $X_0=1$. 于是

$$\begin{aligned} & P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1) \\ & = P\left\{ \sum_{k=1}^{X_n} \xi_{n,k} = j \mid X_n = i, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1 \right\} \\ & = P\left\{ \sum_{k=1}^{i} \xi_{n,k} = j \mid X_n = i, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1 \right\}. \end{aligned}$$

由于 $X_n, \cdots, X_1$ 被 $\{\xi_{n-1,k}\}, \cdots, \{\xi_{1,k}\}$ 决定,而后者与 $\{\xi_{n,k}: k=1,2,\cdots\}$ 相互独立,可见上式

$$\begin{aligned} & P\left\{ \sum_{k=1}^{i} \xi_{n,k} = j \mid X_n = i, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1 \right\} \\ & = P\left\{ \sum_{k=1}^{i} \xi_{n,k} = j \right\} \\ & = P\left\{ \sum_{k=1}^{X_n} \xi_{n,k} = j \mid X_n = i \right\} \triangleq p_{ij}, \end{aligned}$$

其中

$$p_{ij} = P\left\{ \sum_{k=1}^{i} \xi_{n,k} = j \right\} = P\left\{ \sum_{k=1}^{i} \xi_{1,k} = j \right\}$$

与 $n$ 无关. 因而 $\{X_n; n=0,1,\cdots\}$ 是以 $(p_{ij})$ 为转移阵的时齐马氏链.

现在让我们来具体计算 $(p_{ij})$. 注意到 $\{\xi_{n,k}: k=1,2,\cdots\}$ 是独立同分布的随机变量列,考虑母函数:

$$F(x) = \sum_{k=0}^{\infty} f_k x^k. \tag{4.3}$$

由幂级数的乘法规则我们有

$$(F(x))^i = \sum_{k=0}^{\infty} f_k^{(i)} x^k,$$

其中

$$f_k^{(1)} = f_k,$$

$$f_n^{(r+1)} = \sum_{k=0}^{n} f_k^{(r)} f_{n-k}. \tag{4.4}$$

这是因为

$$F(x)^{r+1} = \left(\sum_{k=0}^{\infty} f_k^{(r)} x^k\right)\left(\sum_{m=0}^{\infty} f_m x^m\right)$$

$$= \sum_{n=0}^{\infty}\left(\sum_{k+m=n} f_k^{(r)} f_m\right) x^n$$

$$= \sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} f_k^{(r)} f_{n-k}\right) x^n,$$

这里 $f_n^{(r)}$ 的概率意义是：$P(\xi_{1,1}+\cdots+\xi_{1,r}=n)$. 事实上我们可用归纳法来说明它.

设 $f_n^{(r)}=P(\xi_{1,1}+\cdots+\xi_{1,r}=n)$，则由独立性及(4.4)

$$P(\xi_{1,1}+\xi_{1,2}+\cdots+\xi_{1,r+1}=n)$$

$$=P((\xi_{1,1}+\cdots+\xi_{1,r})+\xi_{1,r+1}=n)$$

$$=\sum_{k=0}^{n} P(\xi_{1,1}+\cdots+\xi_{1,r}=k)P(\xi_{1,r+1}=n-k)$$

$$=\sum_{k=0}^{n} f_k^{(r)} f_{n-k} = f_n^{(r+1)}.$$

此外，我们还可以简单地按下式表出 $p_{ij}$：

$$p_{ij} = f_j^{(i)} = \frac{1}{j!}\frac{\partial^j}{\partial x^j}(F(x))^i\Big|_{x=0} \quad (i \geqslant 1, j \geqslant 0),$$

$$p_{0,0} = 1, \quad p_{0j} = 0\ (\text{当 } j \neq 0). \tag{4.5}$$

最后这行意味着 0 状态（无粒子）是一个吸收态，当过程达到 0 状态就不再有粒子了，分裂也就中止了.

## 2. 消亡与爆炸的概率

设 $\rho_{i,0}$ 为开始是 $i$ 个粒子最终全部消亡的概率，即

$$\rho_{i,0} = P_i(\tau_0(\omega) < +\infty).$$

由于各粒子之间的分裂过程是相互独立的，所以

$$\rho_{i,0} = (\rho_{1,0})^i, \tag{4.6}$$

将 $\rho_{1,0}$简单记为 $\rho$，则 $\rho_{i,0}=\rho^i$.

现在我们来分析分别在什么条件下分别有：$\rho=1$，$\rho=0$ 及 $0<\rho<1$，也就是什么条件下粒子概率为 1 地全死光、永远不会死光、或既可能死光也可能不死光.

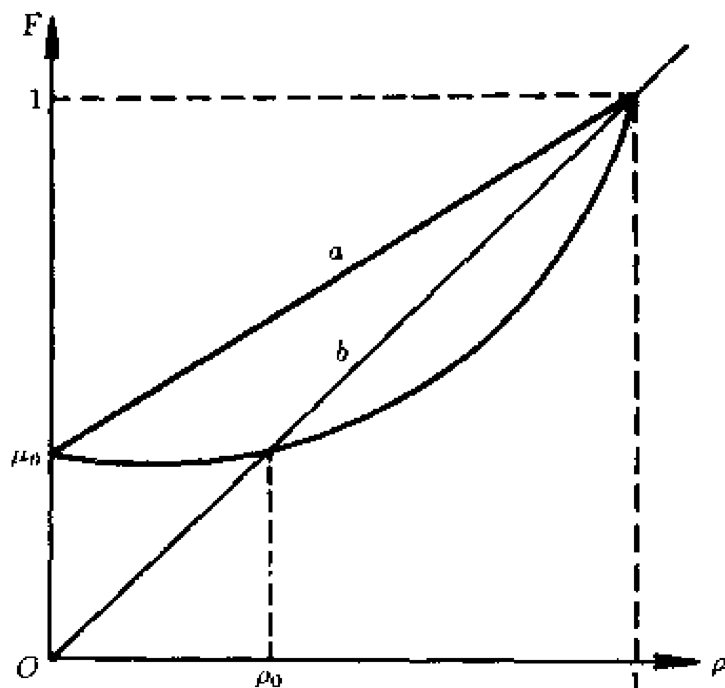
最简单的情况是 $f_0=0$ 与 $f_0=1$，它们分别是一个粒子永远不死与一个粒子在下一个时刻必然死亡，那么显然前者永不会死光，后者是必定死光. 于是为避免这种平凡的情况我们设 $0<f_0<1$.

下面我们来求 $\rho$. 我们记平均一个粒子的后代个数为：

$$\mu \triangleq E\xi_{n,k} = \sum_{j=0}^{\infty} jP(\xi_{n,k}=j) = \sum_{j=1}^{\infty} jf_j. \tag{4.7}$$

直观地猜想：$\mu>1$ 大概不会全死光，$\mu<1$ 大概要全死光.

注意到利用全概公式及(4.6)可推得




$a: \mu = F'(1) < 1$;  $b: \mu = F'(1) > 1$

图 4.1

$$
\begin{aligned}
\rho &= P_1(\tau_0(\omega) < +\infty) \\
&= p_{10} + \sum_{k=1}^{\infty} P_1(X_1 = k) P_k(\tau_0(\omega) < +\infty) \\
&= p_{10} + \sum_{k=1}^{\infty} f_k \rho_{k,0} = f_0 \cdot 1 + \sum_{k=1}^{\infty} f_k \rho_{k,0} \\
&= \sum_{k=0}^{\infty} f_k \rho_{k,0} (\rho_{0,0} = 1) \\
&= \sum_{k=0}^{\infty} f_k \rho^k = F(\rho).
\end{aligned}
$$

也就是说$\rho$应为下面二曲线之交点的横坐标：

$$
y = x \quad 与 \quad y = F(x) \quad (0 \leqslant x \leqslant 1). \tag{4.8}
$$

再注意到$\sum_{k=0}^{\infty} f_k = 1$，即$F(1)=1$，方程(4.8)永远有一个平凡解$x=1$. 又由于当$0<x<1$时，$F'(x) = \sum_{n=1}^{\infty} n f_n x^{n-1} > 0$，而且

$$
y'' = F''(x) = \sum_{n=2}^{\infty} n(n-1) f_n x^{n-2} \geqslant 0 \quad (\forall\ 0 \leqslant x \leqslant 1),
$$

可见$F(x)$是增的凸函数. 于是(4.8)是否在区间(0,1)中有解，就归结为

$$
F'(1) - 1 = (F(x) - x)'|_{x=1}
$$

是正还是负. 前者说明当$\varepsilon>0$充分小时，$F(1-\varepsilon)-(1-\varepsilon)<F(1)-1=0$，因而由

$$
F(0) - 0 = f_0 > 0
$$

与连续函数的中值公式立即得到(4.8)在(0,1)中有解，进而由$F(\cdot)$的凸性就知道(4.8)在(0,1)中解唯一，我们记它为$\rho_0$. 另一方面，在$F'(1)-1\leqslant 0$(即$f\leqslant 1$)时，由$F(x)$的凸性，若$F(x)\neq x$，则立即得到$F(x)-x<0(\forall\ x\in(0,1))$. 这时只可能有一个解

$$
\rho = 1,
$$

从而得知$\rho_{i,0}=\rho^i=1$，这意味着从$i$个粒子出发最终概率为1地全

部灭绝.

在 $\mu>1$ 时,$\rho$ 可能是 1 或(4.8)的另一解 $\rho_0\in(0,1)$. 现在我们来说明在这个情况下,一个粒子的灭绝概率一定是 $\rho_0$. 事实上,我们有

$$\rho_{1,0}=\sum_{n=1}^{\infty}f_{1,0}(n)=\lim_{n\to\infty}\sum_{k=1}^{n}f_{1,0}(k)p_{0,0}(n-k)=\lim p_{1,0}(n).$$

下面我们归纳证明 $p_{1,0}(n)\leqslant\rho_0$,从而就可得到 $\rho=\rho_{1,0}\leqslant\rho_0<1$,即 $\rho=\rho_0$. 为此注意

$$p_{1,0}(1)=f_0=F(0)<F(\rho_0)=\rho_0,$$

也就是当 $n=1$ 时下式成立:

$$p_{1,0}(n)\leqslant\rho_0. \tag{4.9}$$

今设(4.9)对 $n$ 成立,由 $F$ 的递增性我们有

$$\begin{aligned}p_{1,0}(n+1)&=\sum_{k=0}^{\infty}p_{1,k}(1)p_{k,0}(n)=\sum_{k=0}^{\infty}f_kp_{k,0}(n)\\&=\sum_{k=0}^{\infty}f_k(p_{1,0}(n))^k=F(p_{1,0}(n))\\&\leqslant F(\rho_0)=\rho_0.\end{aligned}$$

由归纳法原理我们就得到(4.9)对一切 $n\geqslant1$ 成立. 从而

$$\rho=\rho_{1,0}\leqslant\rho_0<1,$$

可见 $\rho$ 只能取值 $\rho_0$.

综上所述,我们有以下定理:

**定理 4.1** 对于一个 GW 简单分枝过程,设一个粒子的裂变分布为

$$P(\omega:\xi(\omega)=k)=f_k,\ f_1<1\quad(k=0,1,\cdots).$$

令 $\mu\triangleq E\xi(\omega)=\sum_{k=1}^{\infty}kf_k$,则灭绝概率(设开始时有 $i$ 个粒子)为:

$$\rho_{i,0}=\rho^i,$$

其中

$$\rho = \rho_{1,0} = \begin{cases} 1, & \text{当 } \mu \leqslant 1; \\ \rho_0, & \text{当 } \mu > 1. \end{cases}$$

而当 $\mu>1$ 时 $\rho_0$ 为方程(4.8)在(0,1)中的唯一解.

**3. 平均第 $n$ 代的粒子个数(规模)**

由 $i$ 个粒子出发在第 $n$ 代时的平均粒子数为:

$$\zeta_n^{(i)} \triangleq E_i(X_n) = i(E_1 X_n) = i\zeta_n^{(1)}.$$

将 $\zeta_n^{(1)}$ 简单记为 $\zeta_n$, 我们有以下递推公式:

$$\zeta_{n+1} = \sum_{k=1}^{\infty} k p_{1,k}(n+1) = \sum_{k=1}^{\infty} k \sum_{j=1}^{\infty} p_{1,j} p_{j,k}(n)$$

$$= \sum_{j=1}^{\infty} p_{1,j} \zeta_n^{(j)} = \sum_{j=1}^{\infty} j p_{1,j} \zeta_n^{(1)} = \mu\zeta_n = \cdots = \mu^{n+1}.$$

从而 $\zeta_n^{(i)} = i\mu^n$. 由此可见:

当 $\mu<1$ 时, 平均粒子数逐代单调下降趋于 0; 当 $\mu=1$ 时, 各代粒子平均相同; 当 $\mu>1$ 时, 平均粒子数按指数阶逐代上升至无穷.

## §4.2 优化的模拟退火方法

**1. 模拟退火方法的基本想法**

优化问题都可以归结为求某函数(目标函数)在一定范围内的最小值. 但是许多常用的方法往往都是运用某种局部的比较去决定优化过程的发展, 以逐步达到最小值. 例如梯度下降法, 就是每次将目标函数在当时所考察的点与此点的邻域中的点的取值加以比较, 梯度的方向正是指向函数值下降最快的方向, 于是我们就将考察的点沿函数下降最快的方向移动, 希望经过多次这样的移动而达到目标函数最小的点. 但这个算法的实际结果是停止在梯度为零的点上, 这意味着在算到一个目标函数的局部极小值时, 函数

在考察点的值小于它周围领域中的所有点的值,搜索过程就停止了. 我们知道任何一个局部极小值点都满足上述条件,但局部极小值的函数值可能实际上比函数最小值大得多,例如在图 4.2 中从 $x_0$ 出发为用梯度下降法,则搜索过程就会停在 $x_1$,但函数在 $x_1$ 处的值远比最小值大. 于是求得最小值的努力就全寄托于初值的选取.

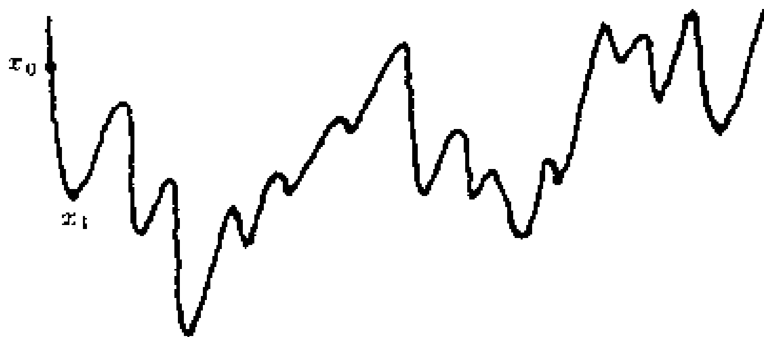



图 4.2

为了克服上述算法的这一问题,Kirkpatrick 等人于 1982 年(见本章附录)提出了模拟退火算法. 它的基本想法是对原决定性的算法引入随机噪声,使得当现时的考察点达到局部极值时,算法过程有一个小概率"逃出"局部极值的陷井. 更具体地说是:在自变量取离散值(可列或有限个值)时,设法将原来算法变成随机的过程——马氏链,使得这个马氏链是一个互通的正常返非周期马氏链,而且其不变分布为

$$\mu_i(\beta) = \exp(-\beta(f(i) - f_0)) \Big/ \sum_i \exp(-\beta(f(i) - f_0)), \tag{4.10}$$

其中 $f(\cdot)$是我们要优化的目标函数,$f_0 = \min_i f(i)$一般未知,而 $\beta$ 是一个控制噪声水平(随机性)的参数. (4.10)式给出的分布很像在物理上的 Boltzmann 分布,其中 $\beta$ 是倒温度,即 $\beta = \frac{1}{T}$($T$ 是绝对温度). 若把无噪声($\beta = +\infty$)的决定性算法看成是绝对零度($T = 1/\beta = 0$)的情形,那么对 $\beta < +\infty$的情形就是将温度升为正温度来进行搜索过程. 注意到在(4.10)式中再令 $\beta \to +\infty$(即 $T \to 0$),

不变分布就全部集中在 $f(\cdot)$ 取最小值的点集上了($f(\cdot)$ 达最小值的点可以不唯一). 这个过程恰似先将温度由绝对零度升高(考虑 $\beta\to+\infty$),再让温度降回绝对零度. 这一升温——降温的过程像是模拟"退火"的过程,因而上述算法又叫模拟退火算法. 即升温以得不变分布,再降温($\beta\to+\infty$)使不变分布集中于 $f$ 的最小值点集上.

**2. 模拟退火算法马氏链的转移阵与理论背景**

我们的基本假定是状态集 $\mathscr{S}$ 为有限集,$f$ 是 $\mathscr{S}$ 到 $R^1$ 的函数. 在前段中我们已经指出,要给出模拟退火算法,关键的一步是要求出一族转移阵 $\{P(\beta)=(p_{ij}(\beta)),\beta_0<\beta<+\infty\}$ 使得(4.10)中的 $\{\mu_i(\beta)\}$ 是 $P(\beta)$ 的不变分布.

例如我们可以取

$$p_{ij}(\beta)=\begin{cases}[1\wedge\exp(-(f(j)-f(i))\beta)]/Z, & i\neq j;\\ 1-\sum\limits_{j\neq i}p_{ij}(\beta), & i=j.\end{cases}$$

(4.11)

其中 $Z$ 是一个充分大的常数,满足 $p_{ij}(\beta)<1$.

例如可取 $Z=\sum\limits_{i,j}[1\wedge\exp(-(f(j)-f(i))\beta)]$ 或(全体状态数)$N$.

在(4.11)中出现的 1 完全是技术性的,它可以使 $Z$ 的计算量不致于太大. 但是当状态集 $\mathscr{S}$ 很大时,以上的取法在计算时实际上是不能实现的,于是我们可用如下的 $P(\beta)$:

$$p_{ij}(\beta)=\begin{cases}G_{ij}(1\wedge\exp(-(f(j)-f(i))\beta))/Z, & i\neq j\\ 1-\sum\limits_{j\neq i}p_{ij}(\beta), & i=j,\end{cases}$$

(4.12)

其中 $(G_{ij})$ 是一个尽量稀疏的互通转移阵,这时 $Z$ 例如可取

$$Z = \sum_{i,j} G_{ij}(1 \wedge \exp(-(f(j) - f(i))\beta)). \qquad (4.13)$$

在 $\mathscr{S}$ 是 $d$-维格点中的有限区域时，$(G_{ij})$ 例如可以取为：

$$G_{ij} = \begin{cases} \frac{1}{|R_i|}, & j \text{ 在 } i \text{ 的邻域 } R_i \text{ 中；} \\ 0, & \text{其他，} \end{cases} \qquad (4.14)$$

其中 $|R_i|$ 表示 $R_i$ 中的点数. 这时 $\beta = +\infty$ 就是从 $i$ 出发以相同的概率向比 $i$ 值小的邻域中的点走. 又如记

$$S_i = \{l \in R_i; f(l) = \min_{j \in R_i} f(j) \text{ 或 } f(l) > f(i)\},$$

并令

$$G_{ij} = \begin{cases} \frac{1}{|S_i|}, & \text{若 } j \in S_i; \\ 0, & \text{其他，} \end{cases} \qquad (4.15)$$

如果 $(G_{ij})$ 还能保持互通，则这样取 $G_{ij}$ 在事实上可以加快计算. 这时当 $\beta = +\infty$ 时它就是决定性优化的最速下降法. 于是这时所得的马氏链 $(P(\beta))$ 可以看成最速下降法的随机摄动.

由于 $(G_{ij})$ 互通及 $1 \wedge \exp(-(f(j) - f(i))\beta) > 0$，可见 $P(\beta)$ 是互通的. 又由于对 $f$ 的严格最小点 $i_0$ 有

$$\begin{aligned} \lim_{\beta \to +\infty} p_{i_0 i_0}(\beta) &= \lim_{\beta \to +\infty} (1 - \sum_{j \neq i_0} p_{i_0 j}(\beta)) \\ &= \lim_{\beta \to +\infty} \left\{ 1 - \sum_{j \neq i_0} G_{i_0 j}[1 \wedge \exp(-(f(j) - f(i_0))\beta)]/Z \right\} \\ &= 1 \quad (\text{因 } f(j) > f(i_0)). \end{aligned}$$

可见 $\beta$ 充分大时 $i_0$ 非周期，因而 $P(\beta)$ 非周期. 记 $\xi^\beta = \{\xi_n^\beta(\omega): n \geqslant 0\}$ 为由(4.12)所定义的 $P(\beta)$ 相对应的马氏链，我们证明当 $(G_{ij})$ 对称时(例如(4.14)中的 $G_{ij}$)它是以(4.10)为可逆不变分布的可逆马氏链. 我们注意当 $f(j) \geqslant f(i)$ 时有(不计常数因子)

$$\begin{aligned} \mu_i(\beta) p_{ij}(\beta) &= e^{-(f(i) - f_0)\beta} (G_{ij} e^{-(f(j) - f(i))\beta})/Z \\ &= G_{ij} e^{-(f(j) - f_0)\beta}/Z \\ &= \mu_j(\beta) p_{ji}(\beta). \end{aligned}$$

同样此等式在 $f(i)\geqslant f(j)$时也应成立，这就证明了我们的断言.

由 $P(\beta)$的不可约性、利用转移矩阵的极限性质我们有

$$P_i^\beta(\omega;\xi_n(\omega)=j)=(P(\beta)^n)_{ij}\rightarrow\mu_j(\beta)\quad(n\rightarrow\infty).$$

令 $S_0\triangleq\{i;f(i)=\min\limits_j f(j)\}$，于是

$$\lim_{\beta\rightarrow\infty}\lim_{n\rightarrow\infty}P_i(\omega;\xi_n(\omega)\in S_0)=1.$$

可见对充分大的 $\beta$，时间充分长后，马氏链几乎总在 $f(\cdot)$最小值点集中.

### 3. 具体算法的说明

在具体计算时我们有两种算法：

a）时齐马氏链的模拟退火算法. 首先任给初值 $x$ 及$\{\beta_n\}$，先对 $\beta_1$ 作马氏链的随机模拟：由 $x$ 出发($\xi_0=x$)，按$\{p_{xj}(\beta_1);j\in\mathscr{S}\}$为分布生成一个随机变量 $\xi_1$ 的样本 $\xi_1(\omega)$，再由$\{p_{\xi_1 j}(\beta_1)\}$生成一个随机变量 $\xi_2$ 的样本 $\xi_2(\omega)$，…重复此算法充分多次，于是 $\xi_n^{(1)}\triangleq\xi_n(\omega)$的分布就近于$\{\mu_j(\beta_1)\}$；再逐个对 $\beta_m$ 重复此步骤得到相应的 $\xi_k^{(m)}(\omega)$，于是这时：

$$\frac{1}{n}\sum_{k=1}^{n}1_j(\xi_k^{(m)}(\omega))\rightarrow\mu_j(\beta_m)\quad(n\rightarrow+\infty).$$

当 $m$ 很大时，除 $S_0$ 中状态外，其他各状态几乎绝少出现，而 $S_0$ 中的状态频繁出现.

b）为了加快计算，在实际算时我们常用非时齐的马氏链，也就是构造第 $m$ 步的转移概率是 $P(\beta_m)$的非时齐马氏链 $\xi_n$ 的轨道，但是使 $\beta_n$ 慢慢地趋于$+\infty$. 例如使 $\beta_n\rightarrow\infty$慢于$\log n$，即：

$$\frac{\beta_n}{\log n}\rightarrow 0\quad(n\rightarrow\infty).$$

可以在理论上证明

$$P(\omega;\xi_n(\omega)\in\mathscr{S}_0)\rightarrow 1\quad(n\rightarrow\infty).$$

于是对 $\omega$ 固定，样本$\{\xi_n(\omega);n\geqslant 0\}$在 $n$ 充分大时的取值就可以认为是 $f(\cdot)$的最小值点.

c) 在用(4.11)或(4.12)来算 $p_{ij}(\beta)$ 及它所生成的马氏链的样本间的转移(为从 $\xi_0$ 转移为 $\xi_1$)时可应用 Gibbs 样本生成法.特别地,此时还可以避免计算 $Z$,只要用可逆性估出 $C_i$(参见(3.36))即可.

d) 模拟退火的计算方法有许多改进与变通,到今仍不失为一个研究课题.我们在本节中只是对其基本方法与思想作了简单的介绍,希望进一步学习的读者可参考以下著作及其中所引文献:

[1] P. J. M. Van Laarhouen & E. H. L. Aarts: Simulated Annealing: Theory and Application, 1987, D. Reidel Publishing Company(可作算法与应用方面的参考).

[2] Hwang, C.-R. & Sheu, S.-J.: Large time behavior of perturbed diffusion Markov process with application to the second eigenvalue problem for Forkker-Plank operators and Simulated annealing, Acta Appl. Math. n, 1990(可作理论方面的参考).

## §4.3 人口结构变化的马氏链模型

### 1. 模型

在研究社会的经济、消费、教育、就业等许多问题时,我们常常会遇到某种分类结构的变化分析.例如考虑社会的教育水平与文化程度的发展变化,我们可以建立以下的模型:首先,我们将某地整个16岁以上的人口分为:文盲、初小、小学、初中、高中、大专、大学、中级技术或专门人才、高级技术与专门人才、特级专家等10级.结构的变化分为升迁或退化(如初小文化者久不接触又变成文盲)和进入与退出(年龄达到16岁或死亡或迁入迁出)两种原因.若用 $(n_1(t),n_2(t),\cdots,n_{10}(t))$ 表示在 $t$ 年各等级的人数, $N(t)=$

$\sum_{i=1}^{10} n_i(t)$ 就是总人数. 若记 $q_{ij}$ 为每年从第 $i$ 级转为第 $j$ 级的人数在 $i$ 级总人数中的百分比，则

$$Q = (q_{ij})_{10\times 10}$$

就是一个准转移阵(即每行加起来不大于1但不一定等于1).

另外我们考虑进入与退出. 记 $\omega_i$ 为每年从第 $i$ 级退出的比例，记 $r_i$ 为每年进入 $i$ 级的人数在总进入人数中的比例. 于是

$$\sum_{j=1}^{10} q_{ij} + \omega_i = 1, \quad r_i \geqslant 0, \quad \sum_{i=1}^{10} r_i = 1. \tag{4.16}$$

又若记 $R(t)$ 为总进入人数，$W(t)$ 为总退出人数，于是

$$N(t+1) = N(t) + R(t) - W(t), \tag{4.17}$$

$$n_j(t+1) = \sum_{i=1}^{10} n_i(t) q_{ij} + r_j R(t).$$

令

$$M(t) = N(t+1) - N(t) = R(t) - W(t). \tag{4.18}$$

若总人数以常数百分比 $\alpha$ 增长($\alpha$ 为负时是减少)，即

$$M(t) = \alpha N(t+1), \qquad \alpha = 1 - \frac{N(t)}{N(t+1)}.$$

于是

$$\frac{n_j(t+1)}{N(t+1)} = \frac{N(t)}{N(t+1)} \sum_{i=1}^{10} \frac{n_i(t)}{N(t)} q_{ij} + r_j \frac{R(t)}{N(t)}$$

$$= \left\{ \sum_{i=1}^{10} \frac{n_i(t)}{N(t)} q_{ij} + r_j \frac{M(t) + W(t)}{N(t)} \right\} (1-\alpha).$$

记 $a_j(t) = \dfrac{n_j(t)}{N(t)}$，上式就改写为

$$a_j(t+1) = \left\{ \sum_{i=1}^{10} a_i(t)(q_{ij} + r_j\omega_i) + \frac{\alpha}{1-\alpha} r_j \right\} (1-\alpha).$$

特别当总人数不变时(即 $\alpha = 0$ 时)就有

$$a_j(t+1) = \sum_{i=1}^{10} a_i(t)(q_{ij} + r_j\omega_i).$$

再记

$$\underset{\sim}{a}(t) = (a_1(t),\cdots,a_{10}(t)), \qquad \tilde{p}_{ij} = q_{ij} + r_j w_i, \qquad \tilde{P} = (\tilde{p}_{ij}),$$

那么上式变为

$$\underset{\sim}{a}(t+1) = \underset{\sim}{a}(t)\tilde{P}. \tag{4.19}$$

这与以 $\tilde{P} = (\tilde{p}_{ij})$ 为转移阵的 Markov 链的 Master 方程

$$\mu_j(i+1) = \sum_i \mu_i(t) p_{ij}$$

是一样的(参见 § 3.1 第二段).

## 2. 稳定的分布结构

在许多问题中,我们希望一个系统能保持一个稳定的分布结构. 从马氏链的不变分布的知识,我们知道若 $\underset{\sim}{a}$ 是 $\tilde{P}$ 的不变分布,那么

$$\underset{\sim}{a} = \underset{\sim}{a}\tilde{P},$$

即从 $\underset{\sim}{a}$ 出发按上面得到的分布的变化规律为:

$$\underset{\sim}{a}(0) = \underset{\sim}{a}, \quad \underset{\sim}{a}(t) = \underset{\sim}{a}$$

这就称 $\underset{\sim}{a}$ 是一个稳定的分布结构.

注意到稳定分布是依赖于进入各级的比例向量 $\underset{\sim}{r}$ 的($\underset{\sim}{r} \triangleq (r_1, r_2,\cdots,r_{10})$). 例如当

$$Q = \begin{pmatrix} 1-a_1 & b_1 & & & & 0 \\ & 1-a_2 & b_2 & & & 0 \\ & & & \ddots & & \\ & & \ddots & & & \\ 0 & & & 1-a_9 & b_9 \\ 0 & & & 0 & 1-a_{10} \end{pmatrix},$$

$$w_i = a_i - b_i \quad (i = 1,\cdots,9), \qquad w_{10} = a_{10}$$

时，若 $\underset{\sim}{r}\triangleq(0,0,\cdots,1)$，则 $\tilde{P}=\begin{pmatrix} & w_1 \\ & \vdots \\ Q & w_8 \\ & a_9 \\ & 1 \end{pmatrix}$，其中 $Q$ 是 $Q$ 的前 9 列. 而

当 $r=(1,0,\cdots,0)$ 时，有

$$P=\begin{pmatrix} 1-b_1 & b_1 & & & & 0 \\ a_2-b_2 & 1-a_2 & b_2 & & & 0 \\ a_3-b_3 & & 1-a_3 & b_3 & & \\ \vdots & & & & \ddots & \\ \vdots & & & \ddots & & \\ a_9-b_9 & & & & 1-a_9 & b_9 \\ a_{10} & & & & 0 & 1-a_{10} \end{pmatrix}.$$

于是 $\underset{\sim}{a}P=\underset{\sim}{a}$ 的解是

$$\underset{\sim}{a}=\left(1,\frac{a}{b_2},\frac{b_1b_2}{a_2a_3},\cdots,\frac{b_1b_2\cdots b_8}{a_2a_3\cdots a_{10}}\right)\Big/Z,$$

其中

$$Z=\sum_{k=1}^{8}\frac{b_1\cdots b_k}{a_2\cdots a_{10}}+1.$$

又例如某地区人口的平均教育水准，规定调入人口的文化程度档必须高于大专. 这时有 $r_i=0(i\leqslant 5)$.

在许多实际问题中，往往 $\underset{\sim}{r}\triangleq(r_1,r_2,\cdots,r_m)$（前面模型中 $m=10$）是可以选择的. 于是，问题就是在于找出 $\underset{\sim}{r}$ 使得它对应的不变分布（稳定结构）比较"合理"（或好）.

例如对人口的年龄结构，往往希望中年人多些. 又例如对人口文化水平结构当然希望文盲越少越好，而专家并不需太多. 于是根据实际问题，我们往往可选择一个合理的稳定结构 $\underset{\sim}{a}^*$，并且从

现在的分布结构 $a(0)$ 出发，让 $\boldsymbol{r}$ 依赖于 $t$，希望逐步调控 $\boldsymbol{r}(t)$ 而尽快达到 $\boldsymbol{a}^*$. 为此我们在这里讨论一下在 $\boldsymbol{r}$ 不依 $t$ 时全部可能的(不同的 $\boldsymbol{r}$ 下)的稳定分布结构. 事实上由于 $\boldsymbol{a}=\boldsymbol{a}\bar{P}$，即

$$\boldsymbol{a}=\boldsymbol{a}(Q+(w_i r_j))=\boldsymbol{a}(Q+(\boldsymbol{w}^T\boldsymbol{r}))=\boldsymbol{a}Q+\boldsymbol{a}\boldsymbol{w}^T\boldsymbol{r},$$

其中 $\boldsymbol{w}=(w_1,w_2,\cdots,w_m)$， $\boldsymbol{r}=(r_1,r_2,\cdots,r_m)$). 那么，当 $\boldsymbol{a}\boldsymbol{w}^T\neq 0$ 时，

$$\boldsymbol{r}=\boldsymbol{a}(I-Q)/\boldsymbol{a}\boldsymbol{w}^T, \tag{4.20}$$

即 $\boldsymbol{a}=\boldsymbol{a}\boldsymbol{w}^T\boldsymbol{r}(I-Q)^{-1}$. 又因为要求 $r_i\geqslant 0$，所以 $a_j\geqslant\sum\limits_i a_i q_{ij}(\forall\ j)$. 这样，对于 $\forall\ \boldsymbol{a}\in\mathscr{A}\triangleq\left\{\boldsymbol{a}:a_j\geqslant\sum\limits_i a_i q_{ij}(\forall\ j)\right\}$ 就可以由(4.20)找出 $\boldsymbol{r}$，它就满足

$$\boldsymbol{a}=\boldsymbol{a}Q+\boldsymbol{a}\boldsymbol{w}^T\boldsymbol{r}=\boldsymbol{a}\bar{P},$$

所以对于这个 $\boldsymbol{r}$，$\boldsymbol{a}$ 是一个稳定结构分布. 可见“合理”的稳定分布必须在 $\mathscr{A}$ 中.

## §4.4 统计力学中的几个常见马氏链模型

### 1. 神经网络的 Hopfield 异步动力学模型

神经网络是一个由 $N$ 个神经元彼此有相互作用而联结在一起的一个体系. 作为神经网络的整个体系协同工作，共同产生一个总的效应. 我们这里考虑一种最简单的神经网络，它是近年来提出的记忆模型——联想记忆的典型.

设每个神经元可以取 $-1$ 与 $+1$ 两个状态，其中 $-1$ 代表该神经元处于抑制态，$+1$ 代表该神经元处于激发态. 于是 $N$ 个神经元的整体可以处于 $2^N$ 个不同的状态，将其全体记为

$$X = \{(x_1, x_2, \cdots, x_N): x_i = -1 \text{ 或 } 1, i = 1,2,\cdots,N\}.$$

我们来给出这种简单神经网络的运动. 当系统处于状态 $\underset{\sim}{x} = (x_1, x_2, \cdots, x_N) \in X$ 时, 如果网络中第 $j$ 个神经元处于激发态时 $x_j = +1$, 则它对于第 $i$ 个神经元有相互作用能 $W_{ij}$, 如果它处于抑制状态 $x_j = -1$, 则它对第 $i$ 个神经元有相互作用能 $-W_{ij}$, 即神经元 $j$ 对 $i$ 的作用是 $W_{ij}x_j$. 于是整个网络对神经元 $i$ 的总相互作用是

$$\sum_{j=1}^{N} W_{ij}x_j \quad (\text{设 } W_{ij} = W_{ji}).$$

在这种相互作用下, 设网络按如下规则运动: 每一步随机地选一个神经元(即每个神经元以 $1/N$ 的机会被选中), 然后看系统对此神经元(设为第 $i$ 个)的总相互作用是否超过一个阈值 $\theta_i$: 若超过 $\theta_i$ 则此步结束时该神经元处于激发状态, 否则它处于抑制态. 由于当系统在现在的状态给定的条件下, 选中哪个神经元以及此神经元在下一步取什么状态都与系统的过去历史无关, 所以若记第 $n$ 步结束时系统所处状态为 $\xi_n$, 则 $\{\xi_n: n = 0,1,2,\cdots\}$ 是一个取值于 $X$ 的马氏链, 其转移阵记为 $P = (p_{\underset{\sim}{x}\underset{\sim}{y}})$, $\underset{\sim}{x}, \underset{\sim}{y} \in X$, 其中

$$p_{\underset{\sim}{x}\underset{\sim}{y}} \triangleq P(\xi_{n+1}^{(\omega)} = \underset{\sim}{y} \mid \xi_n^{(\omega)} = \underset{\sim}{x})$$

$$= \begin{cases} \dfrac{1}{N}, & \text{当 } \underset{\sim}{y} = \underset{\sim}{x}^i \text{ 且 } y_i \neq x_i; \\ 0, & \underset{\sim}{y} \neq \underset{\sim}{x} \text{ 的其他情况;} \\ 1 - \sum\limits_{\underset{\sim}{y} \neq \underset{\sim}{x}} p_{\underset{\sim}{x}\underset{\sim}{y}}, & \underset{\sim}{y} = \underset{\sim}{x}, \end{cases} \quad (4.21)$$

其中

$$\underset{\sim}{x}^i = (x_1, x_2, \cdots, x_{i-1}, y_i, x_{i+1}, \cdots, x_N),$$

$$y_i = \operatorname{sgn}\left(\sum_{j=1}^{N} W_{ij}x_j - \theta_i\right) \quad \left(\text{设} \sum_{j=1}^{N} W_{ij}x_j - \theta_i \text{ 全不为 } 0\right).$$

下面我们记

$$h_i = \left( \sum_{j=1}^{N} W_{ij} x_j - \theta_i \right).$$

**例 4.1** 若 $N=4$,

$$(W_{ij}) = \begin{pmatrix} 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 \end{pmatrix},$$

$$\theta = (\theta_1, \theta_2, \theta_3, \theta_4) = \left( \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2} \right). \qquad (4.22)$$

为符号简单起见,我们用$\frac{1}{2}(\underset{\smile}{x}+\underset{\smile}{1})$所代表的二进制数的十进制数表示作为$\underset{\smile}{x}$的记号(从$\underset{\smile}{x}$变成$\frac{1}{2}(\underset{\smile}{x}+\underset{\smile}{1})$即将所有 $x_i=-1$ 的 $x_i$ 改为 $x_i=0$),也就是$(-1,-1,-1,-1)=0$,$(-1,-1,-1,1)=1$,…,$(1,1,1,-1)=14$,$(1,1,1,1)=15$. 我们可从计算得到下表:

| $\underset{\smile}{x}$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathrm{sgn} h_1$ | − | − | − | + | − | − | − | − | − | + | + | + | − | − | − | + |
| $\mathrm{sgn} h_2$ | + | − | − | − | + | + | + | − | − | − | − | − | + | − | − | − |
| $\mathrm{sgn} h_3$ | − | + | + | + | − | − | − | + | + | + | + | + | − | + | + | + |
| $\mathrm{sgn} h_3$ | − | + | + | + | − | − | − | + | + | + | + | + | − | + | + | + |

注:上表中"+"代表"1","−"代表"−1".

于是我们可以求出 16×16 转移概率阵,例如我们有

$$p_{0,4} = \frac{1}{4}, \qquad p_{0,0} = \frac{3}{4},$$

这是因为从状态 $0=(-1,-1,-1,-1)$出发第一步能选 1,2,3,4 四个神经元之一,而且是等概率的,如选中神经元 1,3 或 4 则由于它们原来处于 0,是抑制态,由上表中状态 0 下方的一列可知第一步后的结果仍为抑制态(等于−1),所以系统此时还未变化,即$p_{0,0}=\frac{3}{4}$;只有选中神经元 2 时,这时的状态由 $0=(-1,-1,-1,$

$-1$)变为 $4=(-1,1,-1,-1)$. 从而 $p_{6,4}=\frac{1}{4}$. 又如

$$p_{14,6}=\frac{1}{4},\qquad p_{7,3}=\frac{1}{4},\qquad p_{7,7}=\frac{3}{4}.$$

$$p_{1,3}=\frac{1}{4},\qquad p_{10,11}=\frac{1}{4},\qquad p_{1,1}=\frac{3}{4}.$$

下面我们算出了这个转移阵的全部分量:$P=(p_{ij})$.

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | $\frac{3}{4}$ | 0 | 0 | 0 | $\frac{1}{4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | $\frac{3}{4}$ | 0 | $\frac{1}{4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | $\frac{3}{4}$ | $\frac{1}{4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | $\frac{3}{4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\frac{1}{4}$ | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | $\frac{1}{4}$ | $\frac{3}{4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | $\frac{1}{4}$ | 0 | $\frac{3}{4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | $\frac{1}{4}$ | 0 | 0 | 0 | $\frac{3}{4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | $\frac{1}{4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | 0 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\frac{3}{4}$ | 0 | $\frac{1}{4}$ | 0 | 0 | 0 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\frac{3}{4}$ | $\frac{1}{4}$ | 0 | 0 | 0 | 0 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 12 | 0 | 0 | 0 | 0 | $\frac{1}{4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\frac{3}{4}$ | 0 | 0 | 0 |
| 13 | 0 | 0 | 0 | 0 | 0 | $\frac{1}{4}$ | 0 | 0 | 0 | $\frac{1}{4}$ | 0 | 0 | 0 | $\frac{1}{4}$ | 0 | $\frac{1}{4}$ |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | $\frac{1}{4}$ | 0 | 0 | 0 | $\frac{1}{4}$ | 0 | 0 | 0 | $\frac{1}{4}$ | $\frac{1}{4}$ |
| 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\frac{1}{4}$ | 0 | 0 | 0 | $\frac{3}{4}$ |

$P$ 的转移阵情况可见图 4.3. 在图中我们可以看到，状态 4 与 11 是两个常返状态，它们各组成一个常返类. $B(11)\triangleq\{1,2,7,3,9,10,15\}$ 是以概率为 1 可达 11 的非常返状态，而 $B(4)\triangleq\{0,5,6,12\}$ 中的状态是以概率为 1 地可达状态 4 的非常返状态. 而非常返状态 8，13，14 分别都是能以正概率可达状态 4 与 11 的非常返状态. 状态 4 与 11 分别为两个常返状态类，都是由一个常返状态组成的一个常返类. 图中可由粗黑线联结的箭头联通的到 4 与 11 的状态分别是以概率为 1 可达 4 与 11 的状态.

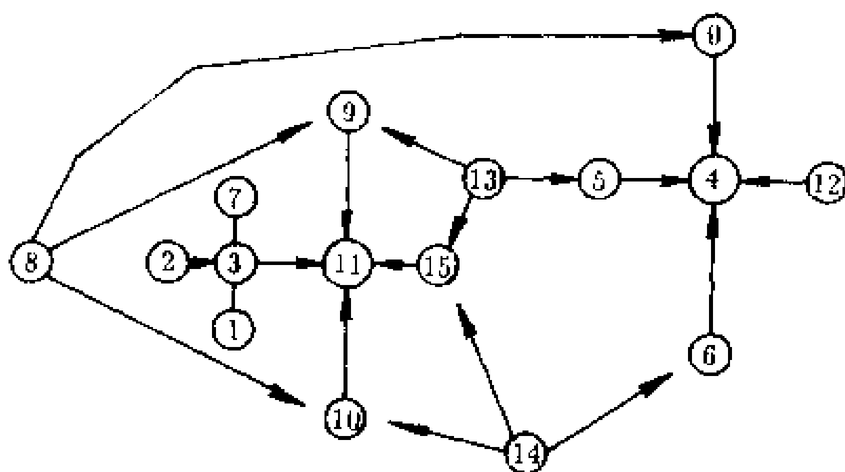



图 4.3

上面这样一个马氏链，可以作为联想记忆系统的一个"玩具模型"：这个系统有两个记忆：即状态 4 与状态 11. $B(11)$ 与 $B(4)$ 中的状态分别可以看成是状态 11 与 4 的部分信息. 这个神经网络系统记住了 $4=(-1,1,-1,-1)$ 与 $11=(1,-1,1,1)$ 这两个状态. 回忆的过程就是从部分信息出发，最终到它的完全信息——系统记住的信息 $(-1,1,-1,-1)$ 或 $(1,-1,1,1)$. 在这个模型下记忆（或称学习）的过程是寻找神经网络的某个适当的相互作用，以保证所要记忆的信息（如 4 和 11）为其常返类中的那些状态. 回忆（或称恢复）所记信息的过程就是从此信息的部分信息（即上面例子中的 $B(11)$ 与 $B(4)$ 中的状态）出发，让系统在规定的相互作用下发展（即让马氏链随 $n$ 发展），当 $n$ 充分大就达到要回忆的信息.

上面讲的记忆方式——联想记忆与传统的记忆方式不同.在传统的记忆方式(如记事本、磁带、磁盘等)中,记入与恢复中的关键是要有一个索引,它指示信息是怎样记录在系统中的,而恢复的过程必须在这个索引的指导下去找出所要的信息.但是这一种记忆的模式与大脑的记忆方式是截然不同的.在大脑的记忆中并不存在什么"索引".而与之相比,联想记忆模式就显得更为接近大脑的记忆模式.所以它是目前被人们普遍认为研究大脑的工作原理与功能的一个较好的模型.

**2. 神经网络的同步动力学随机模型(recurrent network)**

神经网络的同步动力学随机模型与1中讲的Hopfield异步模型很类似,但有两点不同:

1) 在同步模型中不限制每步只改变一个神经元;

2) 这里神经网络的作用并不完全决定神经网络的运动,而只决定它运动的概率分布.也就是说,这里不再随机地选取一个神经元作运动,而是让各个神经元同时作随机运动.具体地说,当系统处于状态 $\underset{\sim}{x}=(x_1,x_2,\cdots,x_N)$时,我们设

$$P_{\underset{\sim}{x},\underset{\sim}{y}}(\beta)\triangleq P(\xi_{n+1}(\xi)=\underset{\sim}{y}\mid\xi_n(\xi)=\underset{\sim}{x})$$

$$=\prod_{i=1}^{N}\frac{1}{1+\exp\left\{-\beta\left(\sum\limits_{j=1}^{N}W_{ij}x_j-\theta_i\right)y_i\right\}}\qquad(4.23)$$

$\left\{\text{这里,我们又设 } h_i(\underset{\sim}{x})=\sum\limits_{j=1}^{N}W_{ij}x_j-\theta_i \text{ 恒不为 } 0\right\}$,其中 $\beta$ 是倒温度参数:$\beta=\frac{1}{T}$,当绝对零度(即 $\beta=+\infty$)时

$$P_{\underset{\sim}{x},\underset{\sim}{y}}(+\infty)=\begin{cases}1, & \underset{\sim}{y}=\underset{\sim}{y}(\underset{\sim}{x});\\ 0, & \text{其他},\end{cases}$$

其中

$$\underset{\sim}{y}(\underset{\sim}{x})=(y_1(\underset{\sim}{x}),\cdots,y_N(\underset{\sim}{x})),\qquad y_i(\underset{\sim}{x})=\operatorname{sgn}h_i(\underset{\sim}{x}).$$

意思是当 $y_i$ 与 $h_i(x)$ 全部同号时 $P(\omega:\xi_{n+1}(\xi)=y(x)|\xi_n(\xi)=x)=1$. 从而当 $\beta$ 很大但有限时，

$$P_{x,y}(\beta)\sim\begin{cases}1, & y=y(x);\\0, & \text{其他}.\end{cases}$$

**例 4.2** 相互作用同例 4.1，由(4.22)给出，这时 $P(+\infty)$ 由 $y=y(x)$ 按(4.23)决定，而 $y(x)$ 由下表给出($\mathrm{sgn}h_i(x)$ 的计算结果参见例 4.1 表)：

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $y(x)$ | 4 | 3 | 3 | 11 | 4 | 4 | 4 | 3 | 3 | 11 | 11 | 11 | 4 | 3 | 3 | 11 |

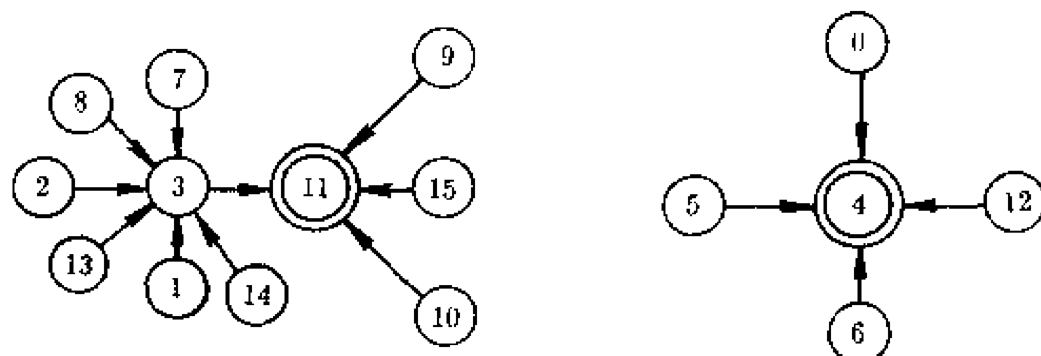



图 4.4

这个神经网络的动力学的转移状况参见图 4.4. 这时系统记住的状态(常返态)仍然是 4 与 11 两个状态，但是它们的部分信息与例 4.1 中的动力学略有不同. 读者可能已经注意到了 $P(+\infty)$ 转移实际上都是以概率为 1 地转移. 所以不妨可理解为：这时相互作用对 $\beta=+\infty$ 所决定的动力学是一个决定性的动力学. 而 $\beta<+\infty$ 的情形是 $\beta=+\infty$ 情形的随机摄动. 当 $\beta$ 充分大时，$\{\xi_n\}$ 以很大的概率按决定性的方式转移，但是此时从状态 4 也可达 11，因为这个马氏链是一个互通的常返马氏链，这样我们就不能用常返互通类来认识系统记住的信息. 注意到对于这个马氏链，当 $\beta$ 充分大，经两步以后，在 $n$ 不是十分大时，如果从 $B(11)\triangleq\{1,2,3,7,8,9,10,11,13,14,15\}$ 出发，$\xi_n$ 以大概率停在 11；如果它从 $B(4)\triangleq\{0,$

5,6,12}出发,$\xi_n$ 以大概率停在 4. 而另一方面,由马氏链的理论可以知道,当 $n$ 趋于无穷时,$\xi_n$ 将跑遍全部状态.

下面我们来证明 $P(\beta)$是可逆的,并求出其可逆不变分布,为简单起见设 $\theta_i=0$, $\theta_i\neq 0$ 的情况可类似地证明.

由于 $y_i$ 与 $\tilde{y}_i$ 只能取−1 或 1,我们有

$$\frac{P_{\underset{\sim}{x},\underset{\sim}{y}}(\beta)}{P_{\underset{\sim}{x},\tilde{\underset{\sim}{y}}}(\beta)}=\frac{\prod_{i=1}^{N}(1+\exp(-\beta h_i(\underset{\sim}{x})\tilde{y}_i))}{\prod_{i=1}^{N}(1+\exp(-\beta h_i(\underset{\sim}{x})y_i))}$$

$$=\prod_{\tilde{y}_i=-y_i}\frac{1+\exp(+\beta h_i(\underset{\sim}{x})y_i)}{1+\exp(-\beta h_i(\underset{\sim}{x})y_i)}=\frac{\prod_{i=1}^{N}(\exp(\beta h_i(\underset{\sim}{x})y_i/2)}{\prod_{i=1}^{N}(\exp(-\beta h_i(\underset{\sim}{x})\tilde{y}_i/2)}$$

$$=\frac{\exp\left(\beta\sum_i h_i(\underset{\sim}{x})y_i/2\right)}{\exp\left(\beta\sum_i h_i(\underset{\sim}{x})\tilde{y}_i/2\right)}.$$

可见

$$P_{\underset{\sim}{x},\underset{\sim}{y}}(\beta)=\frac{\exp\left(\frac{\beta}{2}\sum_i h_i(\underset{\sim}{x})y_i\right)}{Z(\underset{\sim}{x},\beta)},\tag{4.24}$$

其中

$$Z(\underset{\sim}{x},\beta)=\sum_{\underset{\sim}{y}}\exp\left(\frac{\beta}{2}\sum_i h_i(\underset{\sim}{x})y_i\right).\tag{4.25}$$

另一方面,由于 $W_{ij}=W_{ji}$,

$$\sum_i h_i(\underset{\sim}{x})y_i=\sum_{i,j}W_{ij}x_jy_i=\sum_{j,i}W_{ji}x_iy_j$$

$$=\sum_i\left(\sum_j W_{ij}y_j\right)x_i=\sum_i h_i(\underset{\sim}{y})x_i.$$

因而

$$Z(\underset{\sim}{x},\beta)P_{\underset{\sim}{x},\underset{\sim}{y}}(\beta)=\exp\left(\frac{\beta}{2}\sum_i h_i(\underset{\sim}{x})y_i\right)$$

$$=\exp\left\{\frac{\beta}{2}\sum_i h_i(\underset{\sim}{y})x_i\right\}=Z(\underset{\sim}{y},\beta)P_{\underset{\sim}{y},\underset{\sim}{x}}(\beta). \tag{4.26}$$

于是可逆不变分布便是

$$\mu_\beta(\underset{\sim}{x})=\frac{Z(\underset{\sim}{x},\beta)}{Z_\beta}. \tag{4.27}$$

其中

$$Z_\beta=\sum_{\underset{\sim}{x}}Z(\underset{\sim}{x},\beta).$$

我们还可以看到

$$\lim_{\beta\to\infty}\mu_\beta(\underset{\sim}{x})=\lim_{\beta\to\infty}\frac{\sum_{\underset{\sim}{y}}\exp\left\{\frac{\beta}{2}\sum_i h_i(\underset{\sim}{x})y_i\right\}}{\sum_{\underset{\sim}{x}}\sum_{\underset{\sim}{y}}\exp\left\{\frac{\beta}{2}\sum_i h_i(\underset{\sim}{x})y_i\right\}},$$

上式右端极限号下分子的主部是 $\exp\left\{\frac{\beta}{2}\sum_i|h_i(\underset{\sim}{x})|\right\}$，而分母的主部是 $\exp\left\{\frac{\beta}{2}\max_{\underset{\sim}{x}}\sum_i|h_i(\underset{\sim}{x})|\right\}$. 令

$$A=\left\{\underset{\sim}{y}:\sum_i|h_i(\underset{\sim}{y})|=\max_{\underset{\sim}{x}}\sum_i h_i(\underset{\sim}{x})\right\},$$

则

$$\lim_{\beta\to\infty}\sum_{\underset{\sim}{x}\in A}\mu_\beta(\underset{\sim}{x})=1,\quad \lim_{\beta\to\infty}\sum_{\underset{\sim}{x}\notin A}\mu_\beta(\underset{\sim}{x})=0.$$

可见不变分布全部集中在 $\mu_\beta(\underset{\sim}{x})$ 取最大值的状态上.

但是当 $\beta$ 有限时，从记忆模式的观点来看 $P(\beta)$ 与它对应的马氏链当 $n\to\infty$ 时就不合理了：这时无论从什么分布出发最终的分布都是一样的. 这就意味着无论从什么初始的部份信息出发，最终我们以同样的概率得到各个记忆信息——这样的记忆系统就成了“白痴”！事实上，只有在时间不长的时候，我们以大概率达到的状态才是对应于初始部分信息的完全信息. 但是什么叫时间不长？系统怎样来控制？这里存在着严重的问题，解决这些问题的一种办法

是让 $\beta$ 依赖 $n$，并控制 $\beta_n$ 的变化．但是这些讨论已超出了本书的范围，读者可参考：

[1] D. J. Amit  Modeling Brian Function. Cambridge University Press. 1989.

[2] Albeverio, MP Qian & Feng  The role of noise in neural networks. Physical Reoiew E, Vol. 52, No. 6, 6593--6606.

**3. Boltzmann 机**

在前面同步动力学网络中再加进输入神经元与输出神经元，就得到一种更为广泛的神经网络模型，称为 Boltzmann 机（见图）：

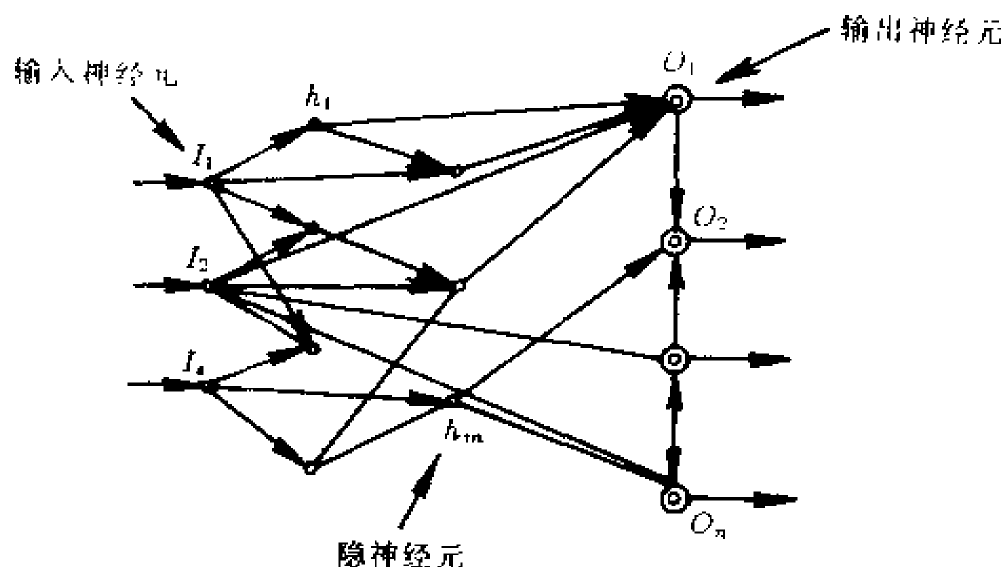



图 4.5

输入神经元的状态由每个时刻的瞬间输入决定，而输出神经元是可观测的．若以 $I_k(t)(k\leqslant s)$、$h_l(t)(l\leqslant m)$，$O_r(t)(r\leqslant n)$ 分别记输入神经元、隐神经元与输出神经元在时刻 $t$ 的值．$I_k(t)$ 的值由即时输入决定，而 $h_l(t+1)$，$O_r(t+1)$ 的值由下列转移概率分别决定：

$$\prod_{k=1}^{m}(1+\exp(-\beta H_k(\underset{\sim}{I}(t),\underset{\sim}{h}(t),\underset{\sim}{O}(t))h_k(t+1))^{-1}$$

$$\times \prod_{r=1}^{n} (1 + \exp(-\beta H_{m+r}(\underset{\smile}{I}(t), \underset{\smile}{h}(t), \underset{\smile}{O}(t)) O_r(t+1))^{-1}.$$

其中($i \leqslant m+n$, $\underset{\smile}{I} = (I_1, \cdots, I_s)$, $\underset{\smile}{h} = (h_1, \cdots, h_m)$, $\underset{\smile}{O} = (O_1, \cdots, O_m)$)

$$H_i(\underset{\smile}{I}, \underset{\smile}{h}, \underset{\smile}{O}) = \sum_{k=1}^{s} W_{ik}^{(I)} I_k + \sum_{l=1}^{m} W_{il}^{(h)} h_l + \sum_{r=1}^{n} W_{ir}^{(O)} O_r.$$

容易看出如果 $\underset{\smile}{I}(t)$ 是一个 Markov 过程，那么($\underset{\smile}{I}(t)$, $\underset{\smile}{h}(t)$, $\underset{\smile}{O}(t)$)也是一个 Markov 过程.

这种网络比同步网络更广，因之也更适合于复杂问题的模式识别. 但是这种网络的学习(学习意即：由样品的实测值来估计相互作用系数($W_{ik}^{(I)}$, $W_{il}^{(h)}$, $W_{ir}^{(O)}$)($i \leqslant m+n$, $k \leqslant s$, $l \leqslant m$, $r \leqslant n$))是一个很复杂的问题. 计算量很大，然而效果比较好，而且很稳健(Robust)(参见[Az]).

另外还有一种重要的神经网络：只允许 $\underset{\smile}{I}$ 对 $\underset{\smile}{h}$ 有作用，$\underset{\smile}{h}$ 对 $\underset{\smile}{O}$ 有作用. 并且对一组输入只运转一次，得到输出 $\underset{\smile}{O}$，这种网络称为随机的前传网络. 当 $\beta = +\infty$ 时就成为一个前传网络. 这种网络在计算机科学、人工智能等领域中有广泛的应用，它是一种比较好的与传统不同的回归模型.

有关神经网络的进一步知识与研究，有兴趣的读者除了可参看前面所引的书[1]外，还可参看

[3] J. Hertz, A. Krogh & R. Polmer　Introduction to the theory of neural computation, LN. 1, Santa Fe Inst. Studies in the Sciences of Complexity. Addison-Wesley Publishing Company, 1991.

## § 4.5　隐 Markov 模型

### 1. 隐马氏模型

在以马氏链作为模型时，必要条件是可以观测到这个链的样

本现实，但是在另一类实际问题中会出现一个马氏链，可是又不能直接观测到这个链的取值样本，而只能观测到这个链的过去取值的某个随机函数（可能还含有干扰或误差），我们由此看出这种模型是由两个随机变量列所组成：一个是不能观测到的有限马氏链，记成$\{S_n:n\geqslant 0\}$，另一个是"与之相关的"可观测到的随机变量列，记成$\{O_n,n\geqslant 0\}$，在这里与之相关的粗略含义是：$\forall\ n$，条件概率

$$P(O_n=v_k\,|\,S_n=i_n,O_{n-1}=v_{k_{n-1}},\cdots,O_1=v_{k_1},S_0=i_0)$$

为已知（最通常的假定是这个条件概率与$(O_{n-1},\cdots,O_1)$的取值无关），于是，$\{S_n:n\geqslant 0\}$称为隐马氏模型，记为HMM. 其中$\{S_n:n\geqslant 0\}$又称为隐马氏链，$\{O_n,n\geqslant 0\}$称为其观测链. 此外，参数$n$也可以是连续的时间参数.

下面举一个例子，以便于读者理解这个模型.

**例 4.3** 有三个盒子内容为：

| | 红球数 | 白球数 |
|---|---|---|
| 盒 1 | 90 | 10 |
| 盒 2 | 50 | 50 |
| 盒 3 | 40 | 60 |

对三个盒分别反复做如下的抽取：

第一盒每次抽取的方式为：随机抽取一球，记下颜色，不放回此球，但放进一个与它颜色不同的球. 第二盒每次抽取的方式为：随机抽取一球记下颜色后放回. 第三盒每次抽取的方式为：随机抽取一球记下颜色不放回，然后放进一个红球. 现在如果某人：随机地选中了一个盒，在此盒中用上述方法得到了一个记录（红，红，红，红，白），但是不告诉我们是哪个盒，我们如何推测他是从哪里抽取的观测样本呢？

令$S_n^{(k)}$＝在第$k$个盒子$(k=1,2,3)$中第$n$次抽取完成后在各盒中的红球数，那么在$k=1,2,3$固定时，$\{S_n^{(k)}:n\geqslant 0\}$各是一个马氏链，其转移阵分别为：

$$P_{i,j}^{(1)} = \begin{cases} \dfrac{i}{100}, & j = i - 1; \\ 1 - \dfrac{i}{100}, & j = i + 1; \\ 0, & \text{其他}, \end{cases}$$

$$P_{i,j}^{(2)} = \begin{cases} 1, & j = i; \\ 0, & j \neq i, \end{cases}$$

$$P_{i,j}^{(3)} = \begin{cases} \dfrac{i}{100}, & j = i; \\ 1 - \dfrac{i}{100}, & j = i + 1. \end{cases}$$

而且

$$S_0^{(1)} = 90, \ S_0^{(2)} = 50, \ S_0^{(3)} = 40.$$

令 $O_n$=记录列中第 $n$ 记录中的白球数($O_n$ 只能取 1 或 0). 于是我们有

$$P^{(1)}\left(O_n = \frac{(S_n - S_{n-1}) + 1}{2} \middle| S_n = i_n, O_{n-1} = v_{k_{n-1}}, S_{n-1} = i_{n-1}, \cdots\right) = 1,$$

$$P^{(2)}(O_n = 0 | S_n = i_n, O_{n-1} = v_{k_{n-1}}, S_{n-1} = i_{n-1}, \cdots)$$

$$= P^{(2)}(O_n = 0) = 0.5,$$

$$P^{(2)}(O_n = 1 | S_n = i_n, O_{n-1} = v_{k_{n-1}}, S_{n-1} = i_{n-1}, \cdots)$$

$$= P^{(2)}(O_n = 1) = 0.5,$$

$$P^{(3)}(O_n = S_n - S_{n-1} | S_n = i_n, O_{n-1} = v_{k_{n-1}}, S_{n-1} = i_{n-1}, \cdots) = 1.$$

由此可算出

$$P^{(1)}((O_1, O_2, O_3, O_4, O_5) = (0,0,0,0,1))$$

$$= P^{(1)}(S_{i+1} = S_i - 1, i = 0,1,2,3, S_5 = S_4 + 1)$$

$$= 0.9 \times 0.89 \times 0.88 \times 0.87 \times 0.13 \approx 0.08,$$

$$P^{(2)}((O_1, O_2, O_3, O_4, O_5) = (0,0,0,0,1))$$

$$= 0.5 \times 0.5 \times 0.5 \times 0.5 \times 0.5 \approx 0.03,$$

$$P^{(3)}((O_1, O_2, O_3, O_4, O_5) = (0,0,0,0,1))$$

$$= 0.4 \times 0.4 \times 0.4 \times 0.4 \times 0.6 \approx 0.015.$$

于是

$P$(该样本列(0,0,0,0,1)是从 $i$ 盒中抽出 | 出现样品(0,0,0,0,1))

$$= \frac{\frac{1}{3}P^{(i)}([(O_1,O_2,\cdots,O_5)=(0,0,0,0,1)]}{\sum_{i=1}^{3}\frac{1}{3}P^{(i)}[(O_1,O_2,\cdots,O_5)=(0,0,0,0,1)]}$$

$$= \begin{cases} 0.64 & i=1; \\ 0.24 & i=2; \\ 0.12 & i=3. \end{cases}$$

可见从第一盒抽出样本(红,红,红,红,白)=(0,0,0,0,1)的概率比从其他两盒中抽出该样本的概率要大得多.从而可以相对有把握地(0.64 的置信度)判断这个样本是从第一盒中抽出来的.同时我们又分别以 0.76 与 0.88 的置信度相信它不是第二盒与第三盒中抽出的.

在这个例子中的不同盒的抽取方式可以抽象为不同的编码方式,从而这个例子启示了用隐马氏模型作模式识别的一种粗略梗概.

隐马氏模型在语音、说话人与视觉信号的模式识别中受到广泛重视,其原因在于这种模型有相当好的"弹性".我们抽象地把一个隐马氏模型记为 $\lambda$,通常它代表一个特定模式,在此模式识别问题中可能出现的不同的隐马氏模型"全体"组成一个"集合"$\Lambda$(全体可能的模式),其中任一个元素 $\lambda$ 为特定的隐马氏模型,其中包括此特定隐马氏模型的转移矩阵与初分布及观测概率$\{P(O_n=v_k|S_n=i_n,S_{n-1}=i_{n-1},\cdots),n\geqslant 1\}$.给定一组样本观测值$(O_1,O_2,\cdots,O_n)$,它与某一个模式(隐马氏模型)$\lambda\in\Lambda$ 的差异程度(或相似程度)可以用 $-\log P(\lambda|O_1,\cdots,O_n)$来度量,它也称为 Bayes 距离(有时也用 $-\log P(O_1,\cdots,O_n|\lambda)$ 来描述$(O_1,\cdots,O_n)$与某一模式的距离),这与 Bayes 的统计思想(更确切地说,假定各模型出现机会均等条件下的 Bayes 推断)相仿,在模式识别中也称为 Bayes 距离

(虽然它并不满足距离公理).

一般对于$(O_1,O_2,\cdots,O_n)$,选取$\lambda^*\in\Lambda$,使$P(O_1,\cdots,O_n|\lambda^*)=\sup\limits_{\lambda\in\Lambda}P(O_1,\cdots,O_n|\lambda)$,则$\lambda^*$是观测$(O_1,O_2,\cdots,O_n)$的最可能模式.

以上的隐马氏模型中,一般地需要$\{P(O_n=v_k|S_n=i_n,O_{n-1}=v_{k_{n-1}},S_{n-1}=i_{n-1},\cdots),n\geqslant 1\}$已知.这个要求太多,包括了过多的参数,在实践上不可能操作,所以在应用中对此又给出了不同的限制,使未知参数尽量少,但又尽量包容有代表性的模型.例如最通常的假定是

$(H_1)$ $P(O_n=v_k|S_n=i,S_{n-1}=i_{n-1},\cdots)=P(O_n=v_k|S_n=i)$
$=P(O_1=v_k|S_1=i)$;

$(H_2)$ $P(S_n=j|O_{n-1}=v_{n-1},S_{n-1}=i_{n-1},O_{n-2}=v_{n-2},\cdots)$
$=P(S_n=j|S_{n-1}=i)=P(S_1=j|S_0=i)$

(即$S.$的值与$O.$的过去值条件独立).满足$(H_1)$,$(H_2)$的典型例子是$O_n=f(S_n)+\eta_n$,其中$f$为已知函数,$\{\eta_n\}$是与马氏链$\{S_n\}$相互独立的随机变量列,代表噪声干扰.记马氏链$\{S_n\}$的初分布为$\underset{\sim}{\pi}$,转移矩阵为$\boldsymbol{P}$,再记

$$b_i(k)=P(O_n=v_k|S_n=i),$$
$$\boldsymbol{B}=(b_i(k)),$$

于是在$(H_1)$,$(H_2)$满足的条件下隐马氏模型的统计性质,即$(O_1,\cdots,O_n,S_1,\cdots,S_n)$的联合分布,完全由参数$\lambda=(\underset{\sim}{\pi},\boldsymbol{P},\boldsymbol{B})$决定,即$(\underset{\sim}{\pi},\boldsymbol{P},\boldsymbol{B})$,它代表了一个特定的隐马氏模型.

## 2. 利用 HMM 作语音识别的简化例子

为了使读者了解 HMM 的作用和优点,我们以 HMM 在语音识别中的应用为例.为此,我们简单化了实际问题,以便突出问题的实质与基本点.此外,尽管在具体应用中细节的研究和处理是十分必要的,但我们还是摆脱了许多细节的考虑,而只指出关键所

在.

一个词的语音可由许多音素组成,例如"发生"这个词由 $F$, $Fa$(表示从 $F$ 到 $a$ 的过渡), $a$, □(空), $sh$, $she$($sh$ 到 $e$ 的过渡), $en$, $n$, $ng$ 等音素组成(后面可加上足够的□(空)),而且每个音素究竟维持多少时间单位,对不同的人,对同一人在不同时间说这个词,都可能有很大不同.于是"发生"这个词的发音就可考虑为一个马氏链,它具有如下形式的转移:

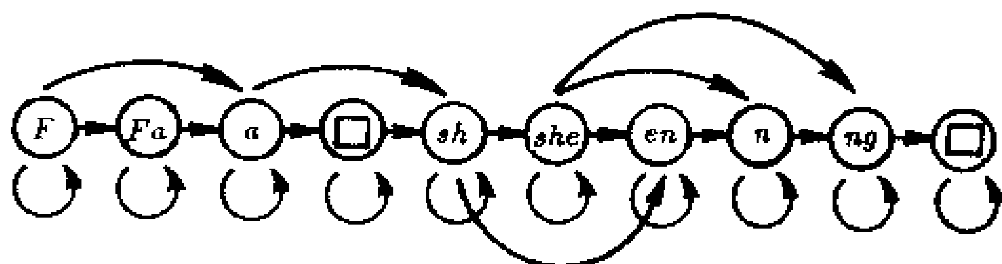



图 4.6

其中状态之间有箭头时表示可能以正概率有转移.实际上,我们听见的并非这个词对应的马氏链,而是声波信号的采样或其在一小段上的 Fourier 变换等经过处理的信号,这就是观测值 $\{O_n, n\geq 1\}$ (这种观测值也可能是连续型随机变量).

对于每个可能接收到的词,我们都构造一个隐马氏链.于是就把语音的机器识别问题化为:从一族(可能是上千个)隐马氏链模型中选取一个与之"最近的"隐马氏链模型,而其代表的词就判定为这个观测对应的词.所以这就是:找 $\lambda^*$ 使

$$P(O_1,\cdots,O_n|\lambda^*) = \sup_{\lambda\in\Lambda_1} P(O_1,\cdots,O_n|\lambda),$$

其中 $\forall\ \lambda(\lambda\in\Lambda_1)$ 代表一个词.

在具体用 HMM 作语音识别时,还需要解决估计统计参数 $(\underset{\sim}{\pi}, \boldsymbol{P}, \boldsymbol{B})$ 的问题(通常称为"学习"或"训练"阶段)以及计算 $P(O_1,\cdots,O_n|\lambda)$ 的问题.通常最常用的是 EM(expectation of maximization)算法,下面我们简述此算法的思想,详细可参考有关专

著，例如：

Lawrence Rabiner &Biing-Hwang Juang, Fundamentals of Speech Recognition, Prentice Hall. Internation Editions, 1993.

L. R. Rabiner, A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition, Proc. IEEE, 77(2), 257--286, 1989.

如何把每个词对应一个 $\lambda=(\pi,\boldsymbol{P},\boldsymbol{B})$ 呢？这就需要这个词的重复发音，把不同 $m$ 个人的同一个发音的声信号的实际 $m$ 个测试值当作 $O_1^{(j)},\cdots,O_n^{(j)}(j\leqslant m)$，由此估计 $\lambda=(\pi,\boldsymbol{P},\boldsymbol{B})$. 在估计 $\lambda$ 时，把 $S_1,S_2,\cdots S_n$ 看成删失的数据，用删失数据的 EM 算法迭代地估计 $\lambda$，其梗概为：用标准人的发音（称为模版）得到初值 $(S_1^{(0)},\cdots,S_n^{(0)})$，由它及 $(O_1^{(j)},\cdots,O_n^{(j)})(j\leqslant m)$ 根据频率估计一个 $\lambda$ 值，记成 $\lambda^{(0)}$，用这个初模型 $\lambda^{(0)}$ 及 $O$（我们用它来代表 $(O_1,\cdots,O_n)$）估计 $S^{(1)}$（代表 $(S_1,\cdots,S_n)$），再用 $O$ 及 $S^{(1)}$ 估计 $\lambda$ 得 $\lambda^{(1)}$，…，归纳地，用 $\lambda^{(k)}$ 及 $O$ 估计 $S^{(k+1)}$，再用 $O$ 及 $S^{(k+1)}$ 估计 $\lambda$ 得 $\lambda^{(k+1)}$，Robbin 证明了：在每次估计时，用适当的修正量，就保证了 $\lambda^{(k)}$ 的收敛性，即存在一个 $\lambda^*$ 使在某种特殊的收敛含义下有 $\lambda^{(k)}\to\lambda^*$，于是模型 $\lambda^*$ 就认为是这个词的模型，从 $O$ 得 $\lambda^*$ 的过程称为"训练"或"学习". 因为计算量相当大，在具体操作中，常把计算简化成一个对 $k$ 的递推算法，这就节约了大量工作量. 再则，如何把一个测量值 $O$ 分成 $n$ 份，成为 $O_1,\cdots,O_n$，也需要视实际情形作不同的探测. 这些细节问题当然不在本书介绍之列.

在具体应用中还可以把一个声母或韵母当作一个隐马氏模型.

### 3. 注

1) $O_n$ 还可以是连续型的随机变量，例如 $(H_1)$ 改为 $P(O_n|S_n,S_n-1,\cdots)$ 为 $N(\mu,\Sigma)$ 或为 $N(\mu_1,\Sigma_1)$，$N(\mu_2,\Sigma_2)$ 的混合分布，这

时 $B$ 变为参数$(\mu,\Sigma)$或$(\mu_1,\mu_2,\Sigma_1,\Sigma_2,a)$($a$ 为混合的系数).

2) 若 $S_n$ 是齐次马氏链,这时它在任一个状态 $i$ 的停留时间 $\tau_i$ 应服从指数分布. 对于有些情况,例如在语音识别中,如果停留时间分布显著的不是指数分布,那样可以改一下模型,设停留时间具有密度 $f(x)$(设它可由几个参数完全确定),转移情况不变,这样的$\{S_n;n\geqslant 1\}$就不再是马氏链了,称为半马氏链. 于是我们的 HMM 改为隐半马氏模型. 可惜的是这种改进模型极不易计算. 又由于 Bayes 距离就是已知样本为 $O$ 的条件下,它在模型 $\lambda$ 中出现的概率. 在实际使用中这两个模型其实差别并不如它看起来那么大. 例如半马氏模型与马氏模型对固定样本 $O$ 得到的 Bayes 距离的差别是很小的(几乎没有). 因之并没有被广泛采用.

3) 数字识别、文字识别也可以用隐马氏模型,只是"时间参数"$n$ 变成了空间参数. 在非在线的文字识别中,往往还需两维的$(n,m)$参数,所以相应的马氏链就应是"双参数马氏链",或者称为马氏场(Markov field).

## §4.6 随机决策模型

设有某种系统,它随机地处在 $S=\{1,2,\cdots,N\}$中的某个状态. 在策略确定为 $a$ 的情况下系统的状态是依照一个矩阵 $P(a)$作为转移概率阵而变化. 这就形成一个决策过程.

让我们先来看一个例子,以便得到一点直观的认识.

**例 4.4** 某经营系统总处在"1","2","3"三种状态之一. 又设在每个整值单位时刻可有两种选择:$d_1$ 或 $d_2$. 假定在采取动作 $d_1$ 和 $d_2$ 时,状态间的转移矩阵分别为

$$P(d_1)=\begin{pmatrix}0.5 & 0.5 & 0\\ 0 & 0.5 & 0.5\\ 0.5 & 0 & 0.5\end{pmatrix},\quad P(d_2)=\begin{pmatrix}0.5 & 0 & 0.5\\ 0.5 & 0.5 & 0\\ 0 & 0.5 & 0.5\end{pmatrix}.$$

又设开始时(即时间 $n=0$ 时)该系统以相等的可能性处在这三个

状态之一，即初始分布为$\left\{\frac{1}{3},\frac{1}{3},\frac{1}{3}\right\}$. 若处在状态 $i$ 并采取 $d_1$ 时能得到报酬为 $g(i,d_1)=2i$，而处在状态 $i$ 并采取 $d_2$ 时相应得到的报酬为 $g(i,d_2)=i^2+0.5$. 我们要在各个时刻有目的地选取动作 $d_1$ 或 $d_2$，使在 $0\leqslant n\leqslant m$ 时的平均累积报酬最大.

假定在时刻 $n$ 采取的动作编号为 $a_n$（它只能是 $d_1$ 或 $d_2$ 之一），令

$$P(a_n)=\begin{cases}P(d_1), & \text{若取 } a_n \text{ 为 } d_1;\\ P(d_2), & \text{若取 } a_n \text{ 为 } d_2.\end{cases}$$

于是由初始分布 $\pi=(\pi_1,\pi_2,\pi_3)=\left\{\frac{1}{3},\frac{1}{3},\frac{1}{3}\right\}$ 及非齐次转移矩阵列 $\{P(a_n)\}$ 决定一个三状态马氏链 $\{\xi_n:n\geqslant 0\}$，这个三状态马氏链 $\xi_n$ 就是在时刻 $n$ 系统所处的状态. 于是在时刻 $m$ 前所得的平均累积报酬为 $E\left\{\sum_{n=0}^{m}g(\xi_n,a_n)\right\}$.

我们简单地限制在时刻 $n$ 所采取的动作只依赖于当时所处的状态，也就是 $a_n$ 只是 $\xi_n$ 的函数 $a_n=f_n(\xi_n)$ ($f_n:\{1,2,3\}\to\{d_1,d_2\}$). 下面我们以 $n=1$ 为例，看看如何求得最高的平均累积报酬. 也就是让 $a_0=f_0(\xi_0)$, $a_1=f_1(\xi_1)$，要选取函数（映射）$f_0,f_1$，使

$$V(f_0,f_1)\triangleq Eg(\xi_0,f_0(\xi_0))+Eg(\xi_1,f_1(\xi_1))=\max.$$

为了易于看到实质（我们把初分布(1/3,1/3,1/3)记为 $\pi=(\pi_1,\pi_2,\pi_3)$），我们注意

$$Eg(\xi_0,f_0(\xi_0))=\sum_i g(i,f_0(i))\pi_i.$$

而因为采取了动作 $a_0=f_0(\xi_0)$，所以马氏链 $\{\xi_n:n\geqslant 0\}$ 从时刻 0 到时刻 1 的转移矩阵应为 $P(f_0(\xi_0))$，于是

$$P(\xi_1=j)=\sum_{i=1}^{3}\pi_i p_{ij}(f_0(i)).$$

从而

$$Eg(\xi_1,f_1(\xi))=\sum_j g(j,f_1(j)P(\xi_1=j)$$

$$= \sum_{i=1}^{3} \sum_{j=1}^{3} g(j, f_1(j)) \pi_i p_{ij}(f_0(i))$$

$$= \sum_{i=1}^{3} \pi_i \sum_{j=1}^{3} g(j, f_1(j)) p_{ij}(f_0(i)) .$$

也就是

$$V(f_0, f_1) = \sum_{i=1}^{3} \pi_i \Big[ g(i, f_0(i)) + \sum_{j=1}^{3} g(j, f_1(j)) p_{ij}(f_0(i)) \Big].$$

从上式可以看出，要选取 $f_0, f_1$ 使 $V(f_0, f_1)$ 最大，就先要选取 $f_1$ 使 $g(j, f_1(j))$ 尽量大. 为此我们注意

$$g(j, d_1) = 2j, g(j, d_2) = j^2 + 0.5;$$

$$(j = 1) \quad g(1, d_1) = 2 > g(1, d_2) = 3/2;$$

$$(j = 2) \quad g(2, d_1) = 4 < g(2, d_2) = 9/2;$$

$$(j = 3) \quad g(3, d_1) = 6 < g(3, d_2) = 19/2.$$

从而可见要使 $g(j, f_1(j))$ 大应取 $f_1 = f_1^*$ 其定义如下：

$$f_1^*(1) = d_1, \quad f_1^*(2) = d_2, \quad f_1^*(3) = d_2.$$

这样就确定了 $f_1$. 记此时的最大报酬为：

$$g^*(j) = g(j, f_1^*(j)) = \begin{cases} 2, & j = 1; \\ 9/2, & j = 2; \\ 19/2, & j = 3. \end{cases}$$

于是 $V(f_0, f_1^*) = \sum_{i=1}^{3} \pi_i \Big[ g(i, f_0(i) + \sum g^*(j) p_{ij}(f_0(i)) \Big]$. 现在看如何选取 $f_0$，使 $V(f_0, f_1^*)$ 最大. 为此只需使方括号中的各量得到最大. 现在与前面不同处只在于现在用方括号中量代替了前面的 $g(j, f_1(j))$. 我们列出它的值，注意到 $p_{ij}(d_1)$ 与 $p_{ij}(d_2)$ 的定义，便得

| | $g(i, d_1) + \sum_{j=1}^{3} g^*(j) p_{ij}(d_1)$ | $g(i, d_2) + \sum g^*(j) p_{ij}(d_2)$ |
|---|---|---|
| $i=1$ | $2+9/2$ | $3/2+19/2$; |
| $i=2$ | $4+19/2$ | $9/2+2$; |
| $i=3$ | $9+2$ | $19/2+9/2$; |

比较上面每一行的大小，可知应取 $f_0=f_0^*$，其定义为：

$$f_0^*(1)=d_2,\quad f_0^*(2)=d_1,\quad f_0^*(3)=d_2.$$

其对应的最大值分别为：11，27/2，14，所以

$$V(f_0^*,f_1^*)=(11+27/2+14)/3=77/6.$$

于是最佳策略为（$f_0^*,f_1^*$ 的定义见上面）：

$$a_0=f_0^*(\xi_0),\quad a_1=f_1^*(\xi_1).$$

而最高平均累积报酬为 $V(f_0^*,f_1^*)=77/6$．

总结一下，从这个例子可以看出要取最佳的 $a_n=f_n(\xi_n)$，即要决定最佳的"状态到动作的对应"$f_n$，而且只要先选定时刻最大的 $m$ 所对应最佳的 $f_m^*$，然后选定最佳的 $f_{m-1}^*$，…，最后选定最佳的 $f_0^*$ 就得到合在一起的最佳行动策略（$f_0^*,\cdots,f_m^*$）．

由这个例子我们可以抽象出下面的数学模型．

设 $P(a)$ 是一个以 $S\triangleq\{1,2,\cdots N\}$ 为状态集合并带有参数 $a$（$\in$ 有限集 $A$）的转移矩阵：

$$P(a)=(p_{ij}(a))_{i,j\leq N},\qquad a\in A.$$

$A$ 称为行动集，它共含有限个行动．

首先我们考虑最简单情形，设在时刻 0，1，…各选一个行动 $a_0,a_1,\cdots(\in A)$，那么由初分布 $\underset{\sim}{\pi}_0$ 及 $\{P(a_n):n\geqslant0\}$ 可决定一个非时齐的马氏链 $\{\xi_n:n\geqslant0\}$，满足：

$$\begin{cases}P(\xi_0=i)=\pi_0(i), & \underset{\sim}{\pi}_0 \text{ 的第 } i \text{ 个分量；}\\ P(\xi_{n+1}=j|\xi_n=i)=p_{ij}(a_n).\end{cases}$$

假定在采取行动列 $\{a_n\}$ 时我们在时刻 $m$ 前能取到的累积报酬为 $\sum_{n=1}^{m}g_n(\xi_n,a_n)$，其中 $g_n(i,a)$ 是在时刻 $n$ 采取行动 $a$ 且这时马氏链 $\xi_n$ 的取值为 $i$ 时能得到的报酬函数．那么，平均报酬为

$$E\sum_{n=1}^{m}g_n(\xi_n,a_n).$$

现在我们考虑较为复杂一些的情形(上面的例子就是属于这种情形)，即 $a_n$ 的取值可依赖于状态 $i$：$a_n = f_n(i)$，其中 $f_n$ 是 $S \to A$ 的映射. 意思是时刻 $n$ 如马氏链的状态为 $i$，则采取决策 $f_n(i)$.

令 $P_n \triangleq (p_{ij}(f_n(i)))_{i,j\in N}$，它仍是一个随机阵. 由$\pi_0$，$\{P_n: n\geqslant 0\}$决定了一个非时齐马氏链$\{\xi_n: n\geqslant 0\}$. 类似地由报酬函数$g_n(i,a)$可以得到时刻 $m$ 前的平均报酬. 由于这个马氏链依赖初分布 $\pi_0$ 及映射列$\{f_n: n\geqslant 0\}$，我们记 $f \triangleq (f_n = f_n(i)), m\geqslant n\geqslant 0)$，并称 $f$ 为一个策略. 记相应的平均报酬为 $E^f_{\pi_0}\{\sum_{n=1}^{m} g_n(\xi_n, a_n)\}$，注意到对于马氏链的一个"轨道"$(\xi_0, \xi_1, \cdots, \xi_m, \cdots)$而言，实际上我们采取的行动列为 $f \triangleq (f_1(\xi_0), f(\xi_1), \cdots, f_m(\xi_m))$(称为一个马氏策略). 这时 $a_n$ 也是随机的，而且 $a_n = f_n(\xi_n)$. 对应于$\pi_0$，$f$ 在时刻 $m$ 前所能取得的平均报酬记为 $J(\pi_0, f)$：

$$J(\pi_0, f) = E^f_{\pi_0}\{\sum g_n(\xi_n, a_n)\}.$$

马氏随机决策的基本问题是：寻找一个马氏策略 $f^*$，使 $J(\pi_0, f^*) = \sup_f J(\pi_0, f)$(这里 sup 取遍所有的马氏策略)，这样的 $f^*$ 如果存在，则称为最佳马氏策略.

若$\forall n$，$f_n$ 不依赖 $\xi_n$，即 $a_n = c_n$(与 $\xi_0, \cdots \xi_n$ 无关的"常"行动)，这种策略记为 $a \triangleq (c_n: m\geqslant n\geqslant 0)$，称为独立常策略. 当然还可以讨论$\{a_n: n\geqslant 0\}$为与$\{\xi_n: n\geqslant 0\}$独立的独立随机列情形，这时 $a = \{a_n: m\geqslant n\geqslant 0\}$称为独立策略.

**定理 4.2** 在所有的马氏策略类中存在最佳马氏策略 $f^*$(未必唯一).

下面的证明过程就是寻找一个最佳马氏策略的具体方法. 为此我们要在马氏策略中构造一个最佳的. 而构造最佳马氏策略 $f^*$ 的方法仿照本节开始的例子中的情形.

取 $f_m^*(i)$，使

$$g_m(i, f_m^*(i)) = \max_a g_m(i, a) (\triangleq g_m^*(i)).$$

再取 $f_{m-1}^*(i)$，使

$$g_{m-1}(i,f_{m-1}^*(i)) + \sum_j p_{ij}(f_{m-1}^*(i))g_m^*(j)$$

$$= \max_a\left[g_{m-1}(i,a) + \sum_j p_{ij}(a)g_m^*(j)\right]\left(\underset{\text{记成}}{=} g_{m-1}^*(i)\right).$$

向后递推地取 $f_k^*(i)$使

$$g_k(i,f_k^*(i)) + \sum_j p_{ij}(f_k^*(i))g_{k+1}^*(j)$$

$$= \max_a\left[g_k(i,a) + \sum_j p_{ij}(a)g_{k+1}^*(j)\right](\triangleq g_k^*(i)).$$

这就得到了$(f_0^*,\cdots,f_m^*)$．

为了证明它是最佳的，只需对$\forall$ $f=(f_n:0\leqslant n\leqslant m)$定义策略：

$$f^{(k)} = (f_0,\cdots f_k,f_{k+1}^*,\cdots,f_m^*),$$

并且用后向数学归纳法直接验证

$$J(\underset{\sim}{\pi_0},f) \leqslant J(\underset{\sim}{\pi_0},f^{(k)}) \ (k \leqslant m).$$

取 $k=0$ 即得 $J(\underset{\sim}{\pi_0},f)\leqslant J(\underset{\sim}{\pi_0},f^*)$，我们略去这段验证.

**注 1**： 以上寻找 $f^*$的办法在计算机上进行时是非常简单的. 问题在于当行动集 $A$ 的元素个数很大时，计算量也会非常大，甚至难以在允许的时间内完成.

**注 2**： 定理 4.2 可以推广到状态集 $S$ 为可列，行动集 $A$ 为紧集，而报酬函数 $g_n(i,a)$为有界连续函数情形.

**注 3**： $m=\infty$的情形：利用有限近似可以证明：如果

$$\sum_{n=0}^{\infty} \sup_{i,a}|g_n(i,a)| < \infty,$$

则存在最佳马氏策略 $f^*$：

$$E_{\pi_0}^{f^*}\left(\sum_{n=0}^{\infty} g_n(\xi_n,a_n)\right) = \sup_f E_{\pi_0}^{f}\left(\sum_{n=0}^{\infty} g_n(\xi_n,a_n)\right).$$

常见例子是折扣报酬模型即折旧模型：

$$g_n(i,a) = r^n g_0(i,a),$$

其中 $0<r<1$ 是折扣因子.

**注 4**： 马氏策略 $f=(f_n;n\geqslant 0)$如满足 $f_n=f_0(\forall\ n)$，则称为平稳马氏策

略．对于折扣报酬模型：$g_n(i,a)=r^n g_0(i,a)$ $(0<r<1)$，记 $J(i,f)\triangleq\sum_{n=0}^{\infty}r^n E_i^f(g_0(\xi_n,a_n))$．设最佳马氏策略为 $f^*=(f_n^*;n\geqslant0)$，再记

$$f^{*(n)}\triangleq(f_0^*,\cdots,f_0^*,f_1^*,f_2^*,\cdots),$$

其中有 $n$ 个 $f_0^*$，那么用归纳法可以证明

$$J(i,f^{*(n)})=J(i,f^{*(n-1)})=\cdots=J(i,f^{*(1)})=J(i,f^*).$$

令 $n\to\infty$，可得到 $J(i,f^{*(\infty)})=J(i,f^*)$．这意味着平稳马氏策略 $f^{*(\infty)}=(f_0^*,f_0^*,\cdots)$ 也是一个最优马氏策略．

利用不动点理论还可以证明：对折扣报酬模型，若 $S$ 可列，$A$ 紧，且 $g_0(i,a)$ 有界，则平稳马氏策略 $f$ 是最佳策略的充要条件是它满足下述 Bellman 方程（动态规划方程）：$\forall\, i\in S$，

$$J(i,f)=\max_a\Big(g_0(i,a)+r\sum_j p_{ij}(a)J(j,f)\Big).$$

此时，$J(\cdot,f)$ 是下述非线性变换 $T(J)$ 的唯一不动点（即 $J(\cdot,f)$ 满足方程 $T(J)=J$）：

$$(T(J))(\cdot)\triangleq\max_a\Big(g_0(\cdot,a)+r\sum_j p_{\cdot j}(a)J(j)\Big).$$

这个结论的优点在于：对于折扣报酬模型，定理 4.2 中的 $m$ 可用 $\infty$ 代替（即近似），而后者的最佳策略是平稳马氏的，只要求出变换 $T$ 的不动点即 $T(J)=J$ 的解 $J(\cdot)$，再令

$$f(i)\triangleq\Big\{a^*:g_0(i,a^*)+r\sum_j p_{ij}(a^*)J(j)=\max\Big\}$$

即可，其计算量将大大地减少．有时用 $T^m(0)$ 来近似 $J(\cdot)$ 也是很有用的，此处 0 是零映射，而 $T^m\triangleq T\circ T\circ\cdots\circ T$ 为映射 $T$ 的 $m$ 次复合．

**注 5：** 最后我们可讨论更为一般的模型：设 $S$ 与 $A$ 仍都为有限集．又设随机变量列 $\{\xi_n\}$ 满足：

$$P(\xi_n=j,\xi_0=i_0,\xi_1=i_1,\cdots,\xi_{n-1}=i_{n-1})=p_{i_{n-1}j}(a_n),$$

其中 $a_n$ 由下式递推地确定：

$$\begin{cases}a_0=f_0(i_0),\\ \cdots\\ a_n=f_n(i_0,a_0,i_1,a_1,\cdots,i_{n-1},a_{n-1},i_n).\end{cases}$$

显见这时 $\{\xi_n;n\geqslant0\}$ 不再是马氏链了．因为它依赖于所有的过去．而且此时 $a_n$ 是一个依赖于 $(\xi_0,\xi_1,\cdots\xi_{n-1})$ 的随机变量．如仍旧记 $f=(f_n;n\geqslant0)$，称 $f$ 为一

个策略. 对于这个模型也类似地可定义最佳策略. 事实上,对于这个模型可以证明:定理 4.2 中的最佳马氏策略也还是这里的最佳策略. 也就是最佳马氏策略在以上定义的非马氏策略类中也仍是最佳的.

## 习　　题

1. 试说明一个家族如世代至多生两个子女,男女有相同的出生率,而只有男性才能继承家族的姓氏,那么此家族的姓氏必将绝传.

2. 在上题中如此家族每代都生三个子女,试求此家族姓氏绝传的概率.

3. (Lamperti 最大分支过程)　令$\{\xi_{n,i};n=0,1,2,\cdots;\quad i=1,2,3,\cdots\}$为独立同分布序列,$P(\xi_{n,i}=j)=\mu_j\quad(j=0,1,\cdots)$,又令:

$$Z_{n+1}=\max\{\xi_{n,1},\cdots,\xi_{n,Z_n}\}.$$

试证明$\{Z_n;n=0,1,\cdots\}$仍是一个马氏链,并求其转移阵. 当

$$\mu_n=\begin{cases}\dfrac{1}{n(n+1)}, & n\geqslant 1;\\ 0, & n=0,\end{cases}$$

证明:

$$\lim_{k\to\infty}P(Z_{n+1}\leqslant kx\mid Z_n=k)=\mathrm{e}^{-x^{-1}}.$$

4. 设简单分支过程的平均分裂个数为$\mu$,$X_0=1$,分裂数的方差为$\sigma^2\left(=\sum_{k=0}^{\infty}(k-\mu)^2\mu_k\right)$,证明

$$\mathrm{Var}X_n=\begin{cases}\sigma^2\mu^{n-1}(\mu^n-1)/(\mu-1), & \text{如 }\mu\neq 1;\\ n\sigma^2, & \text{如 }\mu=1.\end{cases}$$

5. 对以下三种$\mu_i$,给出不灭绝的概率的图解,并求出相继两代间平均规模的递推公式:

a. $\mu_1=0,\mu_0=p,\mu_2=q(p+q=1)$;

b. $\mu_j=qp^j(j=0,1,2,\cdots)$;

c. $p_j = \frac{\lambda^j}{j!}e^{-\lambda}(j=0,1,2,\cdots)$.

6. 生物的外部表征(如肤色、眼睛颜色、头发颜色等)都由一对基因决定,组成这对基因的待选者可以有优势基因(记为 $d$)与劣势基因(记为 $r$)两种,因此这对基因可以是下列之一:优种($dd$)、混种($dr$)、劣种($rr$).后一代的基因对来自随机地选取父本的基因对中的一个与母本的基因对中的一个.

(1) 对任意个体,每次用一混种与之交配,所得的后代仍用混种与之交配,如此继续.请构造一个描写这种传代的马氏链,并求它的不变分布,以及从优种出发首次返回优种的平均代数.

(2) 对任意个体,每次用优种与之交配.请构造马氏链,证明优种是吸收态,并分别求从混种及劣种出发变为优种的平均代数.

7. 色盲的遗传由两种基因 $c$ 与 $s$ 决定,男性只有一个基因($c$ 或 $s$),女性可有两个基因($cc$、$cs$ 或 $ss$),当某人有基因 $c$ 或 $cc$ 时就有色盲.已知关于基因的遗传规律是:男孩只以等概率地继承母亲的两个基因之一,而女孩除了以等概率继承母亲的两个基因之一外,还要继承父亲的那个(可见母亲的色盲必传给男孩,但未必传给女孩,而若父亲为非色盲则女儿一定不是色盲).今设某家有一女儿,其父为色盲,而她又嫁给了一个色盲的男子,他们生下一女;此女又嫁给了一个色盲的男子,又生下第三代女儿,仍嫁给色盲男子,…,如此继续.

试用马氏链描写基因的转移.并说明

$$\lim_{n\to\infty}P(\text{第 } n \text{ 代女孩为色盲})=1.$$

8. 一个服务系统由 $k$ 个工作站串接而成,在服务要求到达第 $j$ 个工作站时能处理的概率为 $p_j$,拒绝处理的概率为 $q_j$,而以 $r_j=1-p_j-q_j$ ($j\leqslant k, r_k=0$)的概率把它转给第 $j+1$ 个工作站处理.试用马氏链来描述,并确定在第 $j$ 个工作站的服务要求平均在经过多少个工作站才得到解决或拒绝处理,被解决与被拒绝的概率各为多大?

# 第五章　Q-过程及其应用

在本章中，我们讨论连续时间参数的马氏链（可数状态空间）及其应用. 在 § 5.1 中我们首先给出最简单的连续时间参数马氏链 — Poisson 过程作为此类马氏过程的典型. 在 § 5.2 中我们进一步讨论较一般的连续时间参数的马氏链及其应用.

## § 5.1　Poisson 过程

Poisson 过程是最简单的连续时间的马氏链. 虽然它简单，但它很有特征地表现了取非负整数值的随机过程的特点，这类过程具有很好的典型性，因而具有广泛的应用，又是许多问题进一步研究的基础. Poisson 过程与 Brown 运动并称随机过程的两大基石.

下面我们就用一个简单例子来说明 Poisson 过程的模型：

### 1. 事故申报的 Poisson 过程模型

**例 5.1**　考虑在时间 $[0,t]$ 中某保险公司的某类事故申报的累计次数为 $\xi_t$，这个数对固定的 $t$ 是随机的，即在时间 $t$ 之前我们并不能确定 $\xi_t$的值. 但经过分析，我们可以认为它具有：

1）**独立增量性**：在互不相交的不同时间区间内事故申报的数量是相互独立的.

2）**时齐性**：在 $(t,t+s]$ 的时间内事故申报的人数的分布与 $t$ 无关而只依赖于 $s$.

3）**普通性**：在无穷小区间 $(t,t+\Delta t]$ 内有两个或两个以上事故申报的概率是 $\Delta t$ 的高阶无穷小，而有限时间区间 $(s,s+t]$ 中的事故申报总是有限的. 且 $\xi_0=0$.

将上面的假设用数学公式表达出来我们就得到下面的译文：

1)′ 对互不相交的区间 $(s_1,t_1],(s_2,t_2],\cdots,(s_m,t_m]$（即 $s_1<t_1\leqslant s_2<t_2\leqslant\cdots\leqslant s_m<t_m$），我们有：相应的随机变量 $\xi_{t_1}-\xi_{s_1},\cdots,\xi_{t_m}-\xi_{s_m}$ 相互独立.

2)′ $\xi_{t+s}-\xi_t$ 对于不同的 $t$ 同分布，于是可令

$$p_k(t)\triangleq P(\xi_{t+s}-\xi_s=k)\quad(k=0,1,2,\cdots)。$$

3)′ $\lim\limits_{\Delta t\to 0}\dfrac{P(\xi_{t+\Delta t}-\xi_t\geqslant 2)}{\Delta t}=0$，$\xi_{t+s}-\xi_t<+\infty(\forall s>0)$.

即

$$p_0(\Delta t)+p_1(\Delta t)=1-o(\Delta t).\tag{5.1}$$

由 1)′ 与 2)′ 进而得到

$$\begin{aligned}p_0(s+t)&=p(\xi_s-\xi_0=0,\xi_{t+s}-\xi_s=0)\\&=p(\xi_s-\xi_0=0)p(\xi_{t+s}-\xi_s=0)\\&=p_0(s)p_0(t).\end{aligned}\tag{5.2}$$

于是对于 $t$ 与整数 $k$，我们有

$$p_0(kt)=(p_0(t))^k,\qquad \left[p_0\left(\frac{t}{k}\right)\right]^k=p_0(t).$$

从而

$$p_0\left(\frac{mt}{k}\right)=(p_0(t))^{\frac{m}{k}},$$

即对于 $\forall t$ 及有理数 $r$ 有

$$p_0(rt)=(p_0(t))^r.\tag{5.3}$$

于是 $(p_0(r))^{\frac{1}{r}}=p_0(1)$，与有理数 $r$ 无关. 若 $p_0(1)=0$，即 $\xi$ 对一切有理数 $r$，$p_0(r)=0$，它蕴含着 $p(\xi_{t+r}-\xi_t\geqslant 1)=1$（$\forall$ 有理数 $r$），也就是说在任意的小区间 $\left[a,\frac{1}{n}+a\right]$ 内至少有一次事故申报. 这样一来在 $(0,1]$ 中就应有无限多次事故申报，因为

$$\begin{aligned}1&\geqslant P(\xi_1-\xi_0=+\infty)=\lim_{n\to\infty}P(\xi_1-\xi_0\geqslant n)\\&\geqslant\lim_{n\to\infty}\prod_{k=1}^{n}P\left(\xi_{\frac{k}{n}}-\xi_{\frac{k-1}{n}}\geqslant 1\right)=1.\end{aligned}$$

这就与 3)中 $\xi_t < +\infty$ 的假定矛盾. 可见 $p_0(1) > 0$. 记 $\log p_0(1) = -\lambda$. 则 $0 \leqslant \lambda < +\infty$, 而且 $p_0(r) = e^{-\lambda r}$ ($r$ 有理). 再利用 $p_0(t) > 0$ 的单调非增性 (若$(0,t]$中无事故, 则$\forall s < t$, $(0,s]$ 中无事故), 由数学分析的常规推理可知

$$p_0(t) = e^{-\lambda t} \quad (\forall\, t). \tag{5.4}$$

再由 1)′～3)′, 我们有

$$\begin{aligned}
P_{k+1}(t+\Delta t) &= P((0,t+\Delta t]\text{ 中恰有 } k+1 \text{ 个事故申报}) \\
&= P((0,t]\text{ 中恰有 } k \text{ 个事故,而}(t,t+\Delta t]\text{ 恰有一次事故}) \\
&\quad + P((0,t]\text{ 中恰有 } k+1 \text{ 个事故,而}(t,t+\Delta t]\text{ 无事故}) \\
&\quad + o(\Delta t) \\
&= p_k(t)p_1(\Delta t) + p_{k+1}(t)p_0(\Delta t) + o(\Delta t) \\
&= p_k(t)(1 - p_0(\Delta t) + o(\Delta t)) + p_{k+1}(t)e^{-\lambda\Delta t},
\end{aligned}$$

$$\begin{aligned}
&\frac{p_{k+1}(t+\Delta t) - p_{k+1}(t)}{\Delta t} \\
&\qquad = p_k(t)\,\frac{1-p_0(\Delta t)}{\Delta t} - p_{k+1}(t)\left(\frac{1-e^{-\lambda\Delta t}}{\Delta t}\right) + o(1).
\end{aligned}$$

令 $\Delta t \to 0$, 右边极限存在, 它是 $p_k(t)\lambda - p_{k+1}(t)\lambda$, 就得到无穷个递推微分方程

$$\frac{dp_{k+1}(t)}{dt} = -\lambda p_{k+1}(t) + \lambda p_k(t) \quad (k = 0,1,2,\cdots). \tag{5.5}$$

由于

$$0 \leqslant p_k(0) \leqslant 1 - p_0(0) = 0 \quad (k \geqslant 1),$$

容易由归纳法计算得到(用常数变异法)

$$p_k(t) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}. \tag{5.6}$$

这说明 $\xi_t$遵从 Poisson 分布, 而

$$\frac{E\xi_t}{t} = \frac{1}{t}\sum_{n=0}^{\infty} n p_n(t) = \frac{\lambda t}{t}\sum_{n=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{n!} = \lambda, \tag{5.7}$$

可见 $\lambda$ 为单位时间内的平均事故率. 而且

$$\frac{1}{t}\mathrm{Var}\xi_t = \frac{1}{t}\sum_{n=0}^{\infty}(n-\lambda t)^2 p_n(t)$$

$$= \frac{1}{t}\sum_{n=0}^{\infty}[n(n-1)+n(1-2\lambda t)+(\lambda t)^2]\left(\frac{(\lambda t)^n}{n!}e^{-\lambda t}\right)$$

$$= \frac{1}{t}\left(\sum_{n=2}^{\infty}\frac{(\lambda t)^{m-2}}{m!}e^{-\lambda t} + \lambda t(1-2\lambda t)+(\lambda t)^2\right)$$

$$= \frac{1}{t}((\lambda t)^2+\lambda t-2(\lambda t)^2+(\lambda t)^2)=\lambda,$$

可见 $\lambda$ 也为单位时间内的平均平方偏差.

**注 1**：3)在 1)与 2)下等价于下述的 $\tilde{3}$)，它在普通书上更常见，而且更易于处理，但是不如 3)直观：

$\tilde{3}$) 在 $(t,t+\Delta t]$ 内有两个或两个以上事故申报的概率为 $o(\Delta t)$，而有一个事故申报的概率与 $\Delta t$ 成正比(这个比率就是 $\lambda$).

**注 2**：生成 $\xi_t$ 的分布列的矩母函数定义为

$$g(t,z) \triangleq \sum_{k=0}^{\infty}p_k(t)z^k,$$

可由(5.4)及(5.5)得到它满足的方程(对 $t$ 的常微分方程)

$$\frac{\partial g(t,z)}{\partial t} = -\lambda g(t,z) + \lambda z g(t,z),$$

及初值

$$g(0,z) = 1.$$

它的解为

$$g(t,z) = e^{\lambda t(z-1)}.$$

由此也可得到 $\xi_t$ 遵从 *Poisson* 分布，这比归纳地解(5.5)能更快地得到结论(5.6)，而且易于推广.

## 2. Poisson 过程作为一个马氏过程

首先，上段的假设 1)说明了随机过程 $\xi=\{\xi_t,t\geqslant 0\}$ 具有独立增量性：即对任意互不相交的时间区间上过程的增量都相互独立. 这类过程一定有马氏性：事实上，对 $0\leqslant s_1<s_2<\cdots<s_m<s$

$< t+s$，以及 $i_1 \leqslant i_2 \leqslant \cdots \leqslant i_m \leqslant i \leqslant j$（因为 *Poisson* 过程是概率为 1 地不会随时间下降的，所以这里取的状态随时间非降）. 考察

$$P(\xi_{t+s}=j|\xi_s=i,\xi_{t_m}=i_m,\cdots,\xi_{t_1}=i_1).$$

它等于

$$P(\xi_{t+s}-\xi_s=j-i|\xi_s=i,\xi_{t_m}=i_m,\cdots,\xi_{t_1}=i_1).$$

由独立增量性及时齐性，上式应为

$$P(\xi_{t+s}-\xi_s=j-i)=P(\xi_t=j-i)=\mathrm{e}^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!}.$$

可见 $\xi=\{\xi_t,t\geqslant 0\}$ 具有这样的统计特性：在已知现在所处状态的条件下，过去与将来发生的事是独立的，也就是说 $\xi$ 具有马氏性. 与第三章中马氏链不同的是，它的时间参数是连续的，所以称为连续参数马氏链. 并且有

$$P(\xi_{t+s}=j|\xi_s=i)=P(\xi_t=j|\xi_0=i).$$

因而它又是时齐的. 它的转移概率可由矩阵表示：

$$(P(\xi_{t+s}=j|\xi_s=i))$$

$$=\begin{pmatrix} 1 & \lambda t & \frac{(\lambda t)^2}{2!} & \frac{(\lambda t)^3}{3!} & \cdots & \frac{(\lambda t)^n}{n!} & \cdots \\ & 1 & \lambda t & \frac{(\lambda t)^2}{2!} & \cdots & \frac{(\lambda t)^{n-1}}{(n-1)!} & \cdots \\ & & \ddots & & & & \\ & & & \ddots & & & \\ & 0 & & & & & \\ & & & & 1 & \lambda t & \cdots \\ & & \ddots & \ddots & & & \end{pmatrix}\mathrm{e}^{-\lambda t}. \tag{5.8}$$

一般地，以 (5.8) 为转移阵且 $\xi_0=0$ 的连续时间参数马氏链叫 Poisson **过程**. 事实上，它等价于说这个随机过程满足第一段中的假设 1)～3).

Poisson 过程是一类非常典型的描述纯随机突发事件的过程。除了上面举出的保险公司的事故申报外，像电话交换台或服务系统的呼唤，放射性粒子的到来，码头泊位船只的到来等等都可以在一定的时间范围内，在一定条件下近似看作 Poisson 过程。

**3. 复合 Poisson 过程(Compound Poisson process)**

若在例 5.1 中各次申报的事故的赔偿费是独立同分布的随机变量列：$Y_1, Y_2, \cdots, Y_n, \cdots$，并与申报次数独立，则到时间 $t$ 保险公司共赔偿的金额为

$$X_t = Y_1 + Y_2 + \cdots + Y_{\xi_t}.$$

设 $Y_k$ 的分布是

$$f(k) = P(Y_k = k) \quad (k = 1, 2, \cdots,).$$

那么当 $j > i$ 时

$$\begin{aligned}
P(X_t = j | X_s = i) &= P(X_t - X_s = j - i | X_s = i) \\
&= P(Y_{\xi_s+1} + Y_{\xi_s+2} + \cdots + Y_{\xi_t} = j - i | Y_1 + \cdots + Y_{\xi_s} = i) \\
&= P(Y_{\xi_s+1} + Y_{\xi_s+2} + \cdots + Y_{\xi_t} = j - i) \\
&= P(Y_1 + Y_2 + \cdots + Y_{\xi_t - \xi_s} = j - i) \\
&= \sum_{k=1}^{\infty} P(Y_1 + Y_2 + \cdots + Y_k = j - i, \xi_t - \xi_s = k) \\
&= \sum_{k=1}^{\infty} P(\xi_{t-s} = k) f^{*k}(j - i) \\
&= e^{-\lambda(t-s)} \sum_{k=0}^{\infty} \frac{(\lambda(t-s))^k}{k!} f^{*k}(j - i).
\end{aligned}$$

其中 $f^{*k}$ 是 $f$ 的 $k$ 重卷积，即

$$f^{*k}(m) \triangleq (f^{*k-1} * f)(m) \triangleq \sum_{l=0}^{m} f^{*k-1}(l) f(m - l).$$

$$f^{*0}(m) \triangleq \begin{cases} 1, & m = 0; \\ 0, & m \geqslant 1. \end{cases}$$

容易看出 $\{X_t;t\geqslant 0\}$ 具有马氏性，而且还是时齐的，其转移矩阵是：

$$(p_{ij}(s,t))=(p_{ij}(t-s)),$$

其中

$$p_{ij}(t-s)=\sum_{k=0}^{\infty}f^{*k}(j-i)\frac{(\lambda(t-s))^k}{k!}e^{-\lambda(t-s)}. \quad (5.9)$$

另一方面，同时对 $s<t$ 我们有

$$P(X_t=i|X_s=i)=P(\xi_t-\xi_s=0)=e^{-\lambda(t-s)}.$$

可见当 $i=j$ 时(5.9)仍然成立.

形如(5.9)的转移矩阵决定的时齐马氏链(其中 $f$ 可以为任何(取整数值的)分布)称为一个取整数值的时齐的**复合 Poisson 过程**(如果赔偿 $Y_i$ 未必取整值，那么对应的 $X_t$ 仍称复合 Poisson 过程，它可以取实数值). 设 $X_0=0$，令 $F(z)=\sum_j f(j)z^j$，则

$$(F(z))^k=\sum_j f^{*k}(j)z^j,$$

进而得到

$$\mu_t\triangleq EX_t=\varphi'(1),$$

$$\sigma_t^2\triangleq EX_t^2-\mu_t^2, EX_t^2=\varphi''(1)+\mu_t,$$

其中

$$\begin{aligned}\varphi(z)&=\sum_{j=0}^{\infty}p_{0,j}(t)z^j=\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}f^{*k}(j)\frac{(\lambda t)^k}{k!}e^{-\lambda t}z^j\\&=\sum_{k=0}^{\infty}\left(\sum_{j=0}^{\infty}f^{*k}(j)z^j\right)\frac{(\lambda t)^k}{k!}e^{-\lambda t}=\sum_{k=0}^{\infty}(F(z))^k\frac{(\lambda t)^k}{k!}e^{-\lambda t}\\&=e^{-\lambda t+F(z)\lambda t}=e^{\lambda t(F(z)-1)}.\end{aligned} \quad (5.10)$$

于是

$$\mu_t=(e^{\lambda t(F(z)-1)})'_{z=1}=\lambda tF'(z)e^{\lambda t(F(z)-1)}\big|_{z=1}=\lambda t\mu,$$

其中

$$\mu=\sum_j f(j)j=EY_1,$$

$$\sigma_t^2=\varphi''(1)+\mu_t-\mu_t^2,$$

$$= \lambda t(F'(1))^2\lambda t + F''(1)) + \mu_t - \mu_t^2$$
$$= \lambda t(\mu^2\lambda t + \sigma^2 + \mu^2 - \mu) + \mu_t - \mu_t^2$$
$$= (\lambda t)^2\mu^2 + \lambda t(\sigma^2 + \mu^2) - \lambda t\mu + \lambda t\mu - (\lambda t\mu)^2$$
$$= \lambda t(\sigma^2 + \mu^2).$$

其中 $\sigma^2 = \mathrm{Var}X_1$. 可见若记 $n = \lambda t$（此即在 $[0,t]$ 中的平均事故申报数）. 那么

$$\text{平均申报总价值} = EX_t = \mu_t = \lambda t\mu = n\mu. \tag{5.11}$$

$$\text{平均申报总价的方差} = E(X_t - EX_t)^2 = \sigma_t^2 = n(\sigma^2 + \mu^2). \tag{5.12}$$

(5.12)是复合 Poisson 过程分布特有的性质. 一般来说，任意 $\{Y_n, n = 1, 2, \cdots\}$ 与 $\xi$ 独立时恒有以下的 Wald 等式(参见 §2.5)：

$$E(Y_1 + Y_2 + \cdots + Y_{\xi_t}) = E(\xi_t)EY_1,$$

但是(5.12)不一定对.

复合 Poisson 过程分布的各种参数的计算可以为我们提供在一定的保险费、保险开支、利率及赔偿的制度下盈利的情况，让各种参数变化制成表格. 就可以供保险公司科学的决策提出可靠的依据. 在更一般的保险问题中，事故申报过程还可以是其他过程，但处理问题的方法与 Poisson 过程极类似，只是这时方差参数就不可能得到像(5.12)那样的显式解，但是我们仍然可用计算机制表得到各种需要的数据.

## §5.2 Q-过程与转移速率阵

### 1. Q 矩阵 —— 转移速率阵

若 $\xi = \{\xi_t; t \geqslant 0\}$ 是连续时间参数的可数状态时齐马氏链，则对 $s_1, s_2, \cdots, s_n < s$，有

$$p_{ij}(t) \triangleq P(\xi_t = j \mid \xi_0 = i)$$

$$= P(\xi_{t+s} = j \mid \xi_s = i, \xi_{s_1} = i_1, \cdots, \xi_{s_n} = i_n).$$

而且,矩阵族 $\{\mathbf{P}(t)\} = (p_{ij}(t); t \geqslant 0)$ 具有以下性质:

$(P_1)$ $0 \leqslant p_{ij}(t) \leqslant 1$, $p_{ij}(0) = \delta_{ij}$. (5.13)

$(P_2)$ $\sum_j p_{ij}(t) = 1$, (5.14)

$(P_3)$ 满足 Kolmogorov 方程:

$$\sum_k p_{ik}(t) p_{kj}(s) = p_{ij}(t+s), \tag{5.15}$$

用矩阵形式写出来就是

$$\boldsymbol{P}(t)\boldsymbol{P}(s) = \boldsymbol{P}(t+s). \tag{5.15$'$}$$

这里$(P_1)$与$(P_2)$和离散的情形完全一样. $(P_3)$与离散情况也类似,只是 $s,t$ 并不一定是非负整数罢了. 与离散时间情形不同的是:在时间离散时 $\boldsymbol{P}(n)$ 可由 $\boldsymbol{P}(1)$ 决定:$\boldsymbol{P}(n) = (\boldsymbol{P}(1))^n$ 这个 $\boldsymbol{P}(1)$ 就决定了全部转移概阵. 但是在 $t$ 连续的情形并不存在一个最小的时间,那么我们是否能对这一情形找到一个相当的量呢?为了求得一些启发,先让我们来看看实数的情况:若有实函数

$$p(n)p(m) = p(n+m) \quad (n,m \text{ 为一切非负整数}),$$

那么一定有 $p(n) = p(1)^n$. 但若有连续函数

$$p(s)p(t) = p(s+t) \quad (\text{对一切非负实数 } s,t \text{ 成立}),$$

那么 $p(t)$ 一定是指数函数:

$$p(t) = ce^{at},$$

其中 $p(0) = c, p'(0) = ca$. 特别若 $p(0) = 1$, 那么 $p(t) = e^{at}$, 其中

$$a = p'(0) = \lim_{t \to 0} \frac{p(t) - p(0)}{t}.$$

由此我们联想到可以考虑

$$\lim_{t \to 0} \frac{p_{ij}(t) - p_{ij}(0)}{t}.$$

例如对 Poisson 过程,这一极限为

$$\lim_{t\to 0}\frac{e^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!}-\delta_{ij}}{t}.$$

可以证明对连续参数的马氏链，一般当 $\lim_{t\to 0}\boldsymbol{P}(t)=I$ 时，上述极限确实有意义，其确切含意是：

$$p_{ij}'(0+)=\lim_{t\to 0}\frac{p_{ij}(t)-p_{ij}(0)}{t}=\lim_{t\to 0}\frac{p_{ij}(t)-\delta_{ij}}{t}\triangleq q_{ij} \tag{5.16}$$

有意义，其中

$$0\leqslant q_{ij}<+\infty \quad (i\neq j), \tag{5.17}$$

$$-\infty\leqslant q_{ii}\leqslant 0, \tag{5.18}$$

而且 $\sum_j q_{ij}\leqslant 0$，即

$$q_i\triangleq -q_{ii}\geqslant\sum_{j\neq i}q_{ij}. \tag{5.19}$$

特别地，若 $\xi$ 是有限状态的马氏链，则 $q_{ij}$ 全部有限，这是因为

$$q_{ii}=p'_{ii}(0+)=\lim_{t\to 0}\frac{p_{ii}(t)-1}{t}=-\lim_{t\to 0}\sum_{j\neq i}\frac{p_{ij}(t)}{t},$$

和号中一共只有有限项，因而和号极限可交换，从而得到

$$q_{ii}=-\sum_{j\neq i}q_{ij}>-\infty.$$

记

$$\boldsymbol{Q}=(q_{ij}).$$

矩阵 $\boldsymbol{Q}$ 到底有什么概率意义呢？由它是否能完全决定 $\{\boldsymbol{P}(t);t\geqslant 0\}$ 呢？

首先考虑状态空间有限的简单情况（设共 $N$ 个状态）由（$P_3$）我们有

$$p'_{ij}(t)=\lim_{\Delta t\to 0}\frac{p_{ij}(t+\Delta t)-p_{ij}(t)}{t}$$

$$=\lim_{\Delta t\to 0}\frac{\sum_{k=1}^{N}p_{ik}(t)p_{kj}(\Delta t)-p_{ik}(t)\delta_{kj}}{t}$$

$$= \sum_{k=1}^{N} p_{ik}(t) \lim_{\Delta t \to 0} \frac{p_{kj}(\Delta t) - \delta_{kj}}{\Delta t}$$

$$= \sum_{k=1}^{N} p_{ik}(t) q_{kj}. \tag{5.20}$$

另一方面

$$p'_{ij}(t) = \lim_{\Delta t \to 0} \frac{p_{ij}(t + \Delta t) - p_{ij}(t)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{\sum_{k=1}^{N} (p_{ik}(\Delta t) - \delta_{ik}) p_{kj}(t)}{\Delta t}$$

$$= \sum_{k=1}^{N} \lim_{\Delta t \to 0} \frac{p_{ik}(\Delta t) - \delta_{ik}}{\Delta t} p_{kj}(t)$$

$$= \sum_{k=1}^{N} q_{ik} p_{kj}(t). \tag{5.21}$$

(5.20)与(5.21)的矩阵形式分别为

$$\boldsymbol{P}'(t) = \boldsymbol{P}(t)\boldsymbol{Q} \tag{5.20$'$}$$

与

$$\boldsymbol{P}(t) = \boldsymbol{Q}\boldsymbol{P}(t), \tag{5.21$'$}$$

它们分别称为 **Kolmogorov 前进方程与后退方程**. 它们是连续时间马氏链的转移阵所可能满足的最重要的方程. 为了确证是否能由 $\boldsymbol{Q}$ 求出 $\boldsymbol{P}(t)$, 我们注意(5.20) 与(5.21)都是 $p_{ij}(t)$ 的线性微分方程, 它们都应满足初始条件:

$$p_{ij}(0) = \delta_{ij} \quad (\boldsymbol{P}(0) = \boldsymbol{I}), \tag{5.22}$$

因此方程(5.20)与(5.21)有唯一解. 令

$$\boldsymbol{P}(t) = \mathrm{e}^{\boldsymbol{Q}t} \triangleq \sum_{n=0}^{\infty} \frac{t^n}{n!} \boldsymbol{Q}^n, \tag{5.23}$$

上式右端级数是收敛的, 事实上

$$q_{ij} \leqslant \max_{1 \leqslant i \leqslant N} \sum_{j \neq i} q_{ij} = \max_{1 \leqslant i \leqslant N} (-q_{ii}) \triangleq q_0.$$

由归纳法易见: $(q_{ij})^n$ 的任意一个元素的绝对值都不超过 $q_0^n$, 于是由

$$\sum_{n=0}^{\infty} \frac{t^n}{n} q_0^n = e^{q_0 t} < +\infty$$

就可得到(5.23)右端任一元素都有限. 再则

$$\sum_{n=0}^{\infty} \frac{t^n}{n!} Q^n = \sum_{n=1}^{\infty} \frac{t^{n-1}}{(n-1)!} Q^n = \left( \sum_{m=0}^{\infty} \frac{t^m}{m!} Q^m \right) Q = Q \sum_{m=0}^{\infty} \frac{t^m}{m!} Q^m.$$

可见

$$(e^{Qt})' = e^{Qt} Q = Q e^{Qt},$$

即 $P(t)$ 与 $e^{Qt}$ 同是(5.20)(或(5.21))满足条件(5.22)的唯一解,因而

$$P(t) = e^{Qt}. \tag{5.24}$$

这正是我们分析实数时希望有的事实.

我们称 $Q$ 为 $\{P(t), t \geqslant 0\}$ 的**转移速率阵**,简称 Q-矩阵.

但是一般来讲,对可数状态马氏链,(5.20)与(5.21)不一定成立,(5.20)在(5.22)下的解也不一定唯一,问题可以十分复杂,附录中对此作了简单介绍.

## 2. Q-矩阵的概率意义及嵌入链

对于一个**轨道右连续**的马氏链(即 $P(\omega;\lim_{t \downarrow t_0} \xi_t(\omega) = \xi_{t_0}(\omega)) = 1$),令 $\tau(\omega) = \inf\{t > 0; \xi_t(\omega) \neq \xi_0(\omega)\}$,这里 $\tau(\omega)$ 是 $\xi_t(\omega)$ 的首次跳跃时间. 我们有下面的定理:

**定理 5.1** 若 $\xi$ 的轨道右连续,而且**保守**$\left( 即对 \forall i 有: \sum_{j \neq i} q_{ij} = -q_{ii} \triangleq q_i < +\infty, 且 q_i > 0 \right)$,则

1) $P(\tau \geqslant t | \xi_0 = i) = e^{-q_i t}$; $P(\tau(\omega) < \infty) = 1$. $\quad$ (5.25)

2) $P(\xi_\tau = j, \tau \leqslant s) | \xi_0 = i) = (1 - e^{-q_i s}) \dfrac{q_{ij}}{q_i}$. $\quad$ (5.26)

3) $P(\xi_\tau = j | \xi_0 = i) = \dfrac{q_{ij}}{q_i} \quad (j \neq i)$. $\quad$ (5.27)

这个定理的意思是当 $\xi$ 处在状态 $i$ 的条件下,它继续留在 $i$ 的

时间是指数分布的,其均值(指数密度)为 $\frac{1}{q_i}$ ,说明 $\xi$ 已在 $i$ 的条件下,当它离开 $i$ 时跳到 $j$ 的概率是 $q_{ij}/q_i$ ,而 1),2)与 3)联系起来说明 $\xi_\tau$ 与 $\tau$ 关于 $\xi_0 = i$ 条件下是条件独立的两个随机变量.

这个定理对认识 Q-矩阵的概率意义很有帮助,但是它们的证明比较繁琐,我们把它略去.有兴趣的读者可参看附录中指出的文献.

下面的定理可以帮助我们进一步认识 Q-矩阵的概率意义,在应用 Q 过程解决实际问题时也很有用处.

**定理 5.2** 设连续时间参数马氏链 $\xi$ 满足定理 5.1 的条件.令 $\tau_k(\omega)$ 是 $\xi$ 的第 $k$ 次跳跃时间.严格地,我们归纳地给出如下定义:

$$\tau_0(\omega) = 0, \tau_1(\omega) = \inf\{t > 0; \xi_t(\omega) \neq \xi_0(\omega)\},$$

$$\tau_{k+1}(\omega) = \inf\{t > \tau_k(\omega); \xi_t(\omega) \neq \xi_{\tau_k}(\omega)\}.$$

记

$$\eta_k(\omega) = \xi_{\tau_k(\omega)}(\omega) \quad (k = 0,1,2,\cdots),$$

则 $\eta \triangleq \{\eta_k(\omega); k = 0,1,\cdots\}$ 是一个与 $\xi$ 有同样状态空间的离散时间参数的时齐马氏链,而且 $\eta$ 的转移矩阵是

$$\boldsymbol{P} = (p_{ij}),$$

其中

$$p_{ij} = \begin{cases} 0, & i = j; \\ \dfrac{q_{ij}}{q_i}, & i \neq j. \end{cases} \tag{5.28}$$

**证示**

$$\begin{aligned} &P(\omega; \eta_{n+1}(\omega) = j | \eta_n(\omega) = i, \eta_{n-1}(\omega) = i_1, \cdots, \eta_0(\omega) = i_n) \\ &\quad = p(\omega; \xi_{\tau_{k+1}(\omega)}(\omega) = j | \xi_{\tau_k(\omega)}(\omega) = i, \\ &\qquad \xi_{\tau_{k-1}(\omega)}(\omega), \cdots, \xi_{\tau_0(\omega)}(\omega) = i_0), \end{aligned} \tag{5.29}$$

其中 $\{\omega; \xi_{\tau_{k+1}(\omega)}(\omega) = j\} = \{\omega; \xi(\omega)$ 在 $\tau_k(\omega)$ 以后首次跳跃时跳到

状态 $j$}.

类似于§3.6中离散时间马氏链的强马氏性，这里对轨道右连续的连续时间马氏链，我们也有强马氏性（证明类似），因而(5.29)式等于

$$P(\omega:\xi_{\tau_1(\omega)}=j|\xi_0(\omega)=i)=\begin{cases}q_{ij}/q_i, & i\neq j;\\ 0, & i=j.\end{cases}$$

上式最后的等号正是(5.27)的结论.

以上讨论说明 $\{\eta_n(\omega):n=0,1,2,\cdots\}$ 的确是一个马氏链，而且其转移阵是 $\boldsymbol{P}=(p_{ij})$，其中

$$p_{ij}=\begin{cases}q_{ij}/q_i, & i\neq j;\\ 0, & i=j.\end{cases}\tag{5.30}$$

$\eta$ 称为 $\xi$ 的**嵌入链**. 它表征了 $\xi$ 在各状态之间的跃迁情况. 若已知 $\eta$，又知道 $q_i(\forall\, i\in$ 状态空间 $\mathscr{S})$，就不仅知道了 $\xi$ 怎样在状态之间跃迁，也知道了在每个状态的停留时间的分布. 也就是说我们就知道了 $\xi$ 的运动的全部统计规律.

## §5.3 几个重要的Q-过程模型

本节给出几个最重要的Q-过程的特例，它们不仅对读者进一步认识Q过程的统计模式及特点有帮助，而且对读者进一步去应用Q-过程也很有启发.

### 1. 连续时间参数的分支过程

设时间 $t=0$ 时有粒子 $X_0$ 个 $(X_0=i_0\geqslant 1)$，每个粒子的分裂时间是一个指数分布的随机变量 $\xi\sim\lambda e^{-\lambda t}I_{[0,\infty)}(t)$，各粒子永远都相互独立，并是完全同样类型的粒子，因而其分裂的概率分布应该相同，设分裂为 $k$ 个的概率是 $f_k(k=0,1,2,\cdots)$. 令 $X_t(\omega)$ 是时刻 $t$ 总共有的粒子数. 下面我们来讨论 $X_t(\omega)$ 变化的统计规律. 类似于考虑 Poisson 过程，我们先考查 $X_t=i$ 时在一个无穷小的区间

$[t, t+h)$ 内，这 $i$ 个独立的粒子中有两个以上粒子分裂的概率是二项分布：

$$\sum_{i \geqslant k \geqslant 2} \binom{i}{k} (e^{-\lambda h})^{i-k}(1 - e^{-\lambda h})^k = O(h^2) = o(h), \quad (5.31)$$

其中 $1 - e^{-\lambda h}$ 与 $e^{-\lambda h}$ 分别是一个粒子在 $[t, t+h)$ 内分裂与不分裂的概率. (5.31)表明有两个以上粒子在长为 $h$ 的时间小区间内分裂的概率只有 $o(h)$，此外，一个粒子在长为 $h$ 的区间内分裂两次以上的概率为：

$$\int_0^h e^{-\lambda s} \lambda (1 - e^{-\lambda(h-s)}) \mathrm{d}s, \quad (5.32)$$

这是因为设第一次分裂的时间是在 $[s, s+\mathrm{d}s]$，在 $s + \mathrm{d}s$ 之后至少分裂一次的概率是$(1 - e^{-\lambda(h-s)})$，由全概率公式就得到(5.32). 容易得(5.32)的如下近似估计：

$$\begin{aligned}
\int_0^h e^{-\lambda s}\lambda(1 - e^{-\lambda(h-s)})\mathrm{d}s &= \int_0^h e^{-\lambda s}\lambda \mathrm{d}s - \lambda e^{-\lambda h}\int_0^h \mathrm{d}s \\
&= 1 - e^{-\lambda h} - \lambda h e^{-\lambda h} = \lambda h - \lambda h e^{-\lambda h} + o(h) \\
&= \lambda h(1 - e^{-\lambda h}) + o(h) = (\lambda h)^2 + o(h) = o(h).
\end{aligned}$$

而一个粒子在 $[t, t+h]$ 中分裂一次的概率为$(1 - e^{-\lambda h}) + o(h)$，由此可见在忽略 $o(h)$ 的意义下，我们只需考虑在 $[t, t+h)$ 中只有一个粒子分裂一次的情形，而其他情形的总概率不过是 $o(h)$. 于是

$$\begin{aligned}
p_{ii}(h) =& P(\text{在}[t, t+h)\text{ 中 } i \text{ 个粒子全不分裂} \mid X_t = i) \\
&+ P(\text{在}[t, t+h)\text{ 中，} i \text{ 个粒子中恰有一个粒子分裂} \\
&\quad \text{一次，分裂结果仍为 1 个粒子} \mid X_t = i) + o(h) \\
=& e^{-\lambda h i} + \binom{i}{1}(1 - e^{-\lambda h})(e^{-\lambda h})^{i-1} \cdot f_1 + o(h) \\
=& e^{-\lambda h i} + i(1 - e^{-\lambda h})f_1 + o(h) = 1 - i\lambda h(1 - f_1) + o(h)
\end{aligned}$$

(请记住 $f_k = P$(一个粒子在分裂时恰分裂成 $k$ 个粒子)). 即

$$q_i = \lim_{h \downarrow 0} \frac{1 - p_{ii}(h)}{h} = \lim_{h \downarrow 0} \frac{i\lambda h(1 - f_1) + o(h)}{h} = i\lambda(1 - f_1).$$

类似地，对 $j \geqslant i+1$，我们有

$$\begin{aligned} p_{ij}(h) &= P(\text{在}[t,t+h)\text{中}, i\text{ 个粒子中恰有一个粒子} \\ &\qquad \text{分裂一次，分裂成 } j-i+1 \text{ 个} \mid X_t = i) + o(h) \\ &= \binom{i}{1}(1-\mathrm{e}^{-\lambda h})(\mathrm{e}^{-\lambda h})^{i-1}\cdot f_{j-i+1} + o(h) \\ &= i\lambda h f_{j-i+1} + o(h), \end{aligned}$$

从而对 $j \geqslant i+1$

$$q_{ij} = \lim_{h\downarrow 0}\frac{p_{ij}(h)}{h} = i\lambda f_{j-i+1}.$$

最后，对 $j < i-1$ 有 $p_{ij}(h) = o(h)$；对 $j = i-1$ 有

$$\begin{aligned} p_{i,i-1}(h) &= P(\text{在}[t,t+h)\text{中恰有一个粒子分裂一次，} \\ &\qquad \text{分裂后消失} \mid X_t = i) + o(h) \\ &= \binom{i}{1}(1-\mathrm{e}^{-\lambda h})(\mathrm{e}^{-\lambda h})^{i-1}\cdot f_0 + o(h) \\ &= i\lambda h f_0 + o(h). \end{aligned}$$

由此有 $q_{i,i-1} = \lim\limits_{h\downarrow 0}\dfrac{i\lambda h f_0 + o(h)}{h} = i\lambda f_0$.

此外，由于现在粒子数确定的条件下，将来的分裂情况与过去独立，这个分枝过程是一个马氏过程，由上面的计算结果知道，它的转移速率阵(Q-矩阵)是 $Q = (q_{ij})$，其中

$$q_{ij} = \begin{cases} i\lambda f_0, & j = i-1; \\ -i\lambda(1-f_1), & j = i; \\ i\lambda f_{j-i+1}, & j \geqslant i+1; \\ 0, & j < i-1. \end{cases} \tag{5.33}$$

## 2. 连续参数的二分枝过程与生灭过程

设在上一段的连续时间参数分支过程中，每个粒子分裂只能有死亡、保持一个或分裂成两个三种情形，即 $f_0 + f_1 + f_2 = 1$，这时可以理解为 $X = \{X_t; t \geqslant 0\}$ 是一个在 0 有吸收壁的生灭过程，

它的转移概率速率阵 $Q=(q_{ij})$ 是：

$$q_{ij}=\begin{cases}0, & i=0;\\ i\lambda f_0, & j=i-1, i\geqslant 1;\\ i\lambda f_2, & j=i+1, i\geqslant 1;\\ -i\lambda(1-f_1), & j=i, i\geqslant 1;\\ 0, & \text{其他}.\end{cases}$$

更一般的生灭过程是指一个马氏链，其转移速率阵为 $Q=(q_{ij})$，其中

$$q_{ij}=\begin{cases}\lambda_i, & j=i-1;\\ \mu_i, & j=i+1;\\ -(\lambda_i+\mu_i), & j=i;\\ 0, & \text{其他}.\end{cases}\tag{5.34}$$

生灭过程相当于连续时间参数的随机徘徊：每次可向前一步或向后跳一步，分别以 $\lambda_i/(\lambda_i+\mu_i)$ 与 $\mu_i/(\lambda_i+\mu_i)$ 的概率从状态 $i$ 向前与向后跳. 从状态 $i$ 出发到起跳的时间服从均值为 $(\lambda_i+\mu_i)^{-1}$ 的指数分布，而且过程的跃迁与它在跃迁前所在状态停留的时间独立.

这里，从一个状态 $i$ 出发到起跳的时间服从指数分布这一特性是过程马氏性的保证. 假若不然，已经知道过程现在(时刻 $t$) 状态 $i$ 的条件下，过程还要经多长时间会起跳的概率分布就要依赖过程在 $t$ 之前已经在 $i$ 停留了多久，过程也就不再有马氏性. 起跳时间服从指数分布时能得到马氏性的实质原因是：若 $\tau$ 服从指数分布($\sim \lambda e^{-\lambda t}$)，则

$$P(\omega:\tau\geqslant t+s\mid\tau\geqslant s)=\frac{p(\omega:\tau(\omega)\geqslant t+s)}{p(\omega:\tau(\omega)\geqslant s)}$$

$$=\frac{\int_{s+t}^{\infty}\lambda e^{-\lambda u}\mathrm{d}u}{\int_{s}^{\infty}\lambda e^{-\lambda u}\mathrm{d}u}=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}=e^{-\lambda t}=P(\omega:\tau\geqslant t).$$

这就意味着当起跳时间服从指数分布时，过程现在处于某状态，此

后经多久从该状态跃迁到其他状态的分布是与过程已在此状态停留多久独立的，因而只要过程跃迁的分布与过去独立，就能保证过程具有马氏性.

**3. $N$ 个服务员的排队系统模型($M/M/N$ 排队模型的排队过程)**

若一个服务系统有 $N$ 个服务员，其顾客到来的规律是一个以 $\alpha$ 为平均密度的 Poisson 过程，每位顾客可由 $N$ 个服务员中任一个不正在服务的服务员接待. 用 $X_t$ 表示时刻 $t$ 时等待服务与正接受服务的所有顾客的总人数，过程 $X$ 称为**排队过程**. 另一方面，当 $X_t$ 大于 $N$ 时就有 $X_t - N$ 个顾客在等待，而 $X_t$ 小于 $N$ 时新到的顾客就立即接受服务. 又设每位顾客的服务时间(从他开始到某空着的服务员处接受服务起计至他离开该服务员为止)服从参数为 $\beta$ 的指数分布，并且不同顾客的服务时间相互独立，还与顾客到来的情况独立. 这时 $\{X_t;t\geqslant 0\}$ 是一个马氏链，因为在 $X_t=i$(说明时刻 $t$ 有 $i$ 个顾客在接受服务或等待)的条件下，$X_{t+s}=j$ 就是在 $[t,t+s)$ 中有 $k$ 个顾客到来，有 $r$ 个顾客结束服务($j=i+k-r$)这两者都与 $t$ 之前的顾客情况独立，也不依赖 $t$. 下面我们来求这个时齐马氏链的 Q- 矩阵，即考查

$$P(\omega:X_{t+h}(\omega)=j|X_t=i)\quad (j=i+k-r).$$

注意到：当 $h\ll 1$ 时

$$P(\omega:\text{在}[t,t+h)\text{ 中有两个以上顾客到来})=o(h),$$

$$P(\omega:\text{在}[t,t+h)\text{ 中有两个以上顾客结束服务})=o(h),$$

$$\begin{aligned}&P(\omega:\text{在}[t,t+h)\text{ 中同时有一个顾客到来，又有一个顾客结束服务})\\&=(1-\mathrm{e}^{-\alpha h})(1-\mathrm{e}^{-\beta h})\\&=(\alpha h)(\beta h)+o(h)=o(h).\end{aligned}$$

我们只要考虑以下三种情况：$[t,t+h)$ 中有一顾客到来，但无顾客结束服务；有一顾客结束服务，但无顾客到来；两者皆无，它们分别与 $p_{i,i+1}(h)$，$p_{i,i-1}(h)$ 与 $p_{ii}(h)$ 只相差 o(h). 令 $l$ 表示 $t$ 时刻正在接受服务的顾客个数，于是

$$l = i \wedge N,$$

$$p_{ii}(h) = \mathrm{e}^{-\alpha h} \cdot \mathrm{e}^{-\beta hl} + o(h) = 1 - (\alpha + \beta l)h + o(h),$$

$$p_{i,i+1}(h) = (1 - \mathrm{e}^{-\alpha h})\mathrm{e}^{-\beta h} + o(h) = \alpha h + o(h),$$

$$p_{i,i-1}(h) = \binom{l}{1}\mathrm{e}^{-\alpha h}(1 - \mathrm{e}^{-\beta h}) + o(h) = l\beta h + o(h),$$

因而对 $i,j \geqslant 0$

$$q_{ij} = \begin{cases} (i \wedge N)\beta, & j = i-1, i \geqslant 1; \\ -(\alpha + (i \wedge N)\beta), & j = i; \\ \alpha, & j = i+1; \\ 0, & \text{其他}. \end{cases} \tag{5.35}$$

容易看出在这个服务模型中，排队过程$\{X_t;t\geqslant 0\}$是上段中的一般生灭过程的一个特例：

$$\lambda_i = (i \wedge N)\beta, \quad \mu_i = \alpha \quad (i = 0,1,\cdots).$$

可以证明这里的 $Q = (q_{ij})$ 能唯一决定一个 Q-过程满足 Kolmogorov 前进方程与后退方程，并具有右连续轨道(参见附录).

## 4. Ising 模型及其 Glauber 动力学

Ising 模型是统计力学中最具典型意义的模型，它在物理、化学、生物等许多领域中都有应用与代表性. 这里我们先简单地介绍有限格点的 Ising 模型. 记 $\mathscr{N} = \{1,2,\cdots,N\}$.

设 $\mathscr{N}^d$($d\geqslant 1$)是每边为 $N$ 个格点的 $d$-维方块中的全部格点集. 每个格点上有一个粒子，它可以处于 $+1$ 与 $-1$ 两种(自旋)状态. 于是整个 $\mathscr{N}^d$ 个粒子组成的系统全部可能的状态集是$\{-1,+1\}^{\mathscr{N}^d}$，将它记为 $\mathscr{S}$，即

$$\mathscr{S} = \{\eta : \eta = (\eta(x)) \in \mathscr{N}^d, \eta(x) \in \{-1,1\}\},$$

对 $\eta\in\mathscr{S}$，称为一个**组态**.

设系统在无穷小的时间区间 $[t,t+h)$($h \ll 1$) 中有两个或两个以上格点改变状态的概率是 $o(h)$. 系统运动的统计规律由下

面的相互作用能量决定：当系统处于状态 $\eta\in\{-1,+1\}^{\Lambda^d}$ 时，粒子系统对于在格点 $x\in\Lambda^d$ 的粒子的"相互作用能量"为

$$H(\eta)=-\frac{1}{2}\sum_{\|x-y\|=1}\eta(x)\eta(y)-h\sum_{x}\eta(x), \qquad (5.36)$$

其中 $\|y-x\|^2=\sum_{i=1}^{d}|y_i-x_i|^2$，即第一和号取遍一切相邻的格点对 $(x,y)$，所谓 $x$ 与 $y$ 相邻是指格点 $x$ 与 $y$ 的 $d$ 个坐标中有且只有一个坐标不同. 而系统从 $\eta$ 在无穷小时间 $h$ 内变为 $\eta^x$（$\eta^x\triangleq$ 将 $\eta$ 中 $\eta(x)$ 变为 $-\eta(x)$ 的组态）的转移概率速率由下式定义

$$\begin{aligned}&\exp(-\beta(H(\eta^x)-H(\eta)))\\&=\exp\{-\beta[\sum_{y:\|x-y\|=1}\eta(x)\eta(y)+h\eta(x)]\},\end{aligned}$$

其中 $\beta$ 是倒温度参数，即它是绝对温度的倒数. 由此决定的 Q-过程就是这个有限格点上的 Ising 模型的 Glauber 动力学. 在§5.5中，我们将看到当 $h=0$ 时这个 Q-过程有可逆平稳分布 $\{\mu(\eta):\eta\in\{-1,+1\}^{\Lambda^d}\}$；这个分布又叫做这个相互作用粒子系统的 **Gibbs 分布**，它与 $\exp\{\beta H(\eta)\}$ 成比例.

## §5.4　Q-过程的极限行为

本节中我们研究 Q-过程的极限行为. 为简单起见，我们这里只讨论保守的（即 $q_i\triangleq-q_{ii}=\sum_{j\neq i}q_{ij}$）轨道右连续的情形，因为对这类过程定理 5.1 与定理 5.2 已经对它们在各状态的停留与跃迁的统计特征用 Q-矩阵给出了清晰的描述.

### 1. 常返、非常返与状态分类

在§3.2中，我们在定义 3.4 中给出的离散时间马氏链常返的定义. 对定义 3.4 稍加修改就适用于连续时间参数相应的情况：

**定义 5.4** 马氏链 $\xi \triangleq \{\xi_t(\omega); t \geqslant 0\}$ 的状态 $i$ 称常返，如果

$$P_i(\omega: \xi(\omega) \text{ 到达又离开 } i \text{ 无穷多次}) = 1;$$

状态 $i$ 称为暂态，如果

$$P_i(\omega: \xi(\omega) \text{ 到达又离开 } i \text{ 无穷多次}) = 0.$$

由于在 $\xi$ 发生无穷多次跳跃前，$\xi$ 在状态之间跃迁的情况与其嵌入链 $\eta$ 是一样的，只要 $\xi$ 在有限时间内不会发生无穷多次跳跃，就保证 $\xi$ 与 $\eta$ 的常返性与互通性相同. 在本章的附录中我们将证明当 $q_i$ 有界，即 $q_i \leqslant c(\forall\ i)$ 时，$\xi$ 在有限时间内概率为 1 地不会发生无穷多次跳跃. 这样一来，$\xi$ 的常返性就可用

$$\dot{\boldsymbol{P}} = (\tilde{p}_{ij}) = (q_{ij}(1 - \delta_{ij})/q_i)$$

及 $\dot{\boldsymbol{P}}^n$，由 §3.2 中有关的定理来认识. 特别地，若 $\xi$ 是有限状态连续参数的马氏链，$(q_{ij}) = \boldsymbol{Q}$ 是互通的矩阵，即 $\dot{\boldsymbol{P}}$ 是互通的，那么 $\xi$ 也是一个互通常返的马氏链.

**例 5.2** （$M/M/N$ 排队过程的常返性）

在 §5.3.3 中引入的 $M/M/N$ 排队过程的 Q-矩阵由(5.35)式给出：

$$q_{ij} = \begin{cases} (i \wedge N)\beta, & j = i - 1, i \geqslant 1; \\ -(\alpha + (i \wedge N)\beta), & j = i, i \geqslant 0; \\ \alpha, & j = i + 1, i \geqslant 0; \\ 0, & \text{其他}. \end{cases}$$

由于 $q_{ij} \leqslant q_i = \alpha + (i \wedge N)\beta \leqslant \alpha + N\beta (j \neq i)$，可见 $\boldsymbol{X}$ 正是满足上面讲到的 $q_i$ 关于 $i$ 有界的条件的. 于是 $\boldsymbol{X}(\omega)$ 在有限时间只能跳跃有限次，从而 $\boldsymbol{X}$ 的互通性与常返性与其嵌入链相同. $\boldsymbol{X}$ 的嵌入链的转移阵是

$$\tilde{p}_{ij} = \begin{cases} \dfrac{(i \wedge N)\beta}{\alpha + (i \wedge N)\beta}, & j = i - 1, i \geqslant 1; \\ \dfrac{\alpha}{\alpha + (i \wedge N)\beta}, & j = i + 1, i \geqslant 0; \\ 0, & \text{其他}. \end{cases}$$

容易看出 $X$ 是互通的，而且在 $i \geqslant N$ 时，$p_{i,i+1} = \frac{\alpha}{\alpha + N\beta}$，$p_{i,i-1} = \frac{N\beta}{\alpha + N\beta}$，它同向右概率为 $\frac{\alpha}{\alpha + N\beta}$ 与向左概率为 $\frac{N\beta}{\alpha + N\beta}$ 的齐次随机徘徊完全相同. 于是从 $i \geqslant N$ 出发若 $\alpha > N\beta$ 则 $\tilde{P}$ 对应的嵌入链 $\eta$ 有正概率从 $i$ 向右走永不回到 $i$，可见 $\eta$ 这时非常返(暂态). 因而 $X$ 也非常返. 当 $\alpha \leqslant N\beta$ 时，从 $i$ 出发向右走，在有限时间内 $\eta$ 以概率为 1 的回到 $i$；而从 $i$ 向左走只有 $(0,1,\cdots,i)$ 有限个状态，如果在 $i$ 设一个吸收壁，这时这个马氏链是一个总共 $i+1$ 个状态(有限状态)互通的以 $i$ 为吸收壁的马氏链，它应以概率为 1 地在有限时间到 $i$，可见马氏链 $\eta$ 是常返的，从而 $X$ 也是常返的.

**例 5.3** （空间齐次的单边生灭过程的常返性）

考虑一种特殊的生灭过程，设 $\xi$ 的 Q-矩阵是 $Q = (q_{ij})$，其中

$$q_{ij} = \begin{cases} \mu, & j = i - 1, i \geqslant 1, \\ \lambda, & j = i + 1, i \geqslant 0, \\ -(\lambda + \mu), & i = j, i \geqslant 1, \\ \lambda, & i = j, i = 0 \end{cases} \qquad (0 < \lambda, \mu < +\infty).$$

这种情况下，由例 5.2 我们也可知道 $\xi$ 在有限时间区间只跳跃有限次，从而 $\xi$ 与其嵌入链 $\eta$ 有相同的互通性与常返性. 由于

$$p_{ij} = \begin{cases} \frac{\mu}{\lambda + \mu}, & j = i - 1, i \geqslant 1, \\ \frac{\lambda}{\lambda + \mu}, & j = i + 1, i \geqslant 0, \\ 1, & i = 0, j = 1, \\ 0, & \text{其他}, \end{cases} \qquad (0 < \lambda, \mu < +\infty).$$

这时嵌入链 $\eta$ 恰为 0 点有反射壁且以 $p = \frac{\lambda}{\lambda + \mu}$ 与 $q = \frac{\mu}{\lambda + \mu}$ 分别为向右与向左概率的简单随机徘徊. 可见当 $\mu < \lambda$ 时 $\xi$ 非常返，$\lambda \leqslant \mu$ 时，$\xi$ 常返.

又若$\xi$是在整个$\mathbf{Z}$(一维整格点)上的空间齐次生灭过程,则与上面的讨论类似我们可以得到当$\lambda \neq \mu$时,$\xi$非常返;当$\lambda = \mu$时,$\xi$常返.

**2. Q-过程转移阵的极限**

为了回避比较深入的概率论与测度论的问题,我们本段中的推导仅限于有限状态($N$个状态)的Q-过程.但其结论仍对$q_i = \sum_{j \neq i} q_{ij}$且对$i$有界的可数状态的情形也适用.

由于

$$\lim_{\delta \to 0} \frac{p_{ii}(\delta) - 1}{\delta} = q_{ii} = -q_i > 0,$$

可见当$\delta$充分小$p_{ii}(\delta) > 0$,因此对充分小的$\delta(\delta \leqslant \delta_0)(p_{ij}(n\delta))$一定是非周期的.设$(q_{ij}/q_i)$是互通的,我们就有$(p_{ij}(\delta))$也是互通的($\delta$充分小).于是对固定的$\delta>0$(充分小)

$$\lim_{n \to \infty} p_{ij}(n\delta) = \pi_j^{\delta}, \tag{5.38}$$

而且对$\forall\ 0 < \delta_1 < \delta_2$充分小有

$$\begin{aligned}
\pi_j^{\delta_2} &= \lim_{n \to \infty} p_{ij}(n\delta_2) = \lim_{n \to \infty} p_{ij}(n(\delta_1 - \delta_2) + n\delta_1) \\
&= \lim_{n \to \infty} \sum_{k=1}^{N} p_{ik}(n(\delta_2 - \delta_1)) p_{kj}(n\delta_1) \\
&= \sum_{k=1}^{N} \pi_k^{\delta_3} \pi_j^{\delta_1} \quad (\delta_3 = \delta_2 - \delta_1) \\
&= \pi_j^{\delta_1}.
\end{aligned}$$

因此只要$\delta$充分小$(\delta < \delta_0)$,(5.38)中的极限与$\delta$无关,即

$$\lim_{n \to \infty} p_{ij}(n\delta) = \pi_j \quad (\forall\ \delta \leqslant \delta_0).$$

又由于任给$\varepsilon > 0$,$\exists\ \delta_1 > 0$,使$\delta < \delta_1$时,对$\forall\ j$

$$|p_{ij}(\delta) - \delta_{ij}| \leqslant 1 - p_{ii}(\delta) \leqslant \varepsilon/2,$$

于是

$$\left|p_{ij}(t)-p_{ij}\left(\left[\frac{t}{\delta_1}\right]\delta_1\right)\right|\leqslant\sum_k p_{ik}\left(\left[\frac{t}{\delta_1}\right]\delta_1\right)|p_{kj}(\delta)-\delta_{kj}|\leqslant\varepsilon/2$$

$\left(其中\ \delta=t-\left[\frac{t}{\delta_1}\right]\delta_1<\delta_1\right)$，可见当 $t\to\infty$

$$\lim_{t\to\infty}p_{ij}(t)=\pi_j,\qquad 即\ P(t)\to\mathbf{1}\underset{\sim}{\pi}.\tag{5.38$'$}$$

这就说明互通的 Q 过程的转移概率阵的极限存在，而且是一个各行都相同的矩阵.

此外，对 $\forall\ t\geqslant 0$，易见

$$\sum_i\pi_i p_{ij}(t)=\sum_i\lim_{s\to\infty}p_{ki}(s)p_{ij}(t)=\lim_{s\to\infty}p_{kj}(t+s)=\pi_j.$$

将上式写成矩阵形式就是

$$\underset{\sim}{\pi}\boldsymbol{P}(t)=\underset{\sim}{\pi},$$

并且有

$$\sum_i\pi_i=1\quad(即\ \underset{\sim}{\pi}\mathbf{1}=1).$$

即 $\underset{\sim}{\pi}$ 具有这样的特点：以它为初分布，$\xi$ 在任何时候的分布都仍为 $\underset{\sim}{\pi}$.

**定义 5.5** 设分布 $\underset{\sim}{\pi}$ 使得对 $\forall\ t\geqslant 0$ 有

$$\underset{\sim}{\pi}\boldsymbol{P}(t)=\underset{\sim}{\pi},\qquad\underset{\sim}{\pi}\mathbf{1}=1,$$

则称 $\underset{\sim}{\pi}$ 是 $\{\boldsymbol{P}(t);t\geqslant 0\}$（或 $\xi$）的（平稳）不变分布.

(5.38)′中给出的 $\underset{\sim}{\pi}$ 正是一个平稳不变分布. 由于通常对 Q-过程我们并不知道 $\{\boldsymbol{P}(t);t\geqslant 0\}$ 是什么，而真正知道的是转移密度（速率）阵 $(\boldsymbol{Q}=(q_{ij}))$，所以我们应设法用 $\boldsymbol{Q}$ 来刻画 $\underset{\sim}{\pi}$. 事实上，由(5.38)′我们有

$$\lim_{\delta\to 0}\sum_i\pi_i\frac{(p_{ij}(\delta)-\delta_{ij})}{\delta}=0,$$

即
$$\sum_i\pi_i q_{ij}=0\quad(\forall\ j).$$

其矩阵形式为

$$\underset{\sim}{\pi} Q = 0, \qquad \underset{\sim}{\pi}\mathbf{1} = 1. \tag{5.38''}$$

可见在 $Q$ 是互通时，$\underset{\sim}{\pi}$ 是(5.38)''的唯一解. 又由于当 $Q$ 互通时，嵌入链 $P$ 是互通的. 因此下面的方程

$$\tilde{\underset{\sim}{\pi}} P = \tilde{\underset{\sim}{\pi}}, \qquad \tilde{\underset{\sim}{\pi}}\mathbf{1} = 1$$

有唯一解，即

$$\sum_{i\neq j}\tilde{\pi}_i q_{ij}/q_i = \tilde{\pi}_j, \quad \tilde{\underset{\sim}{\pi}}\mathbf{1} = 1,$$

也即

$$\pi_i = \tilde{\pi}_i/q_i Z, \tag{5.39}$$

其中 $Z$ 是归一化常数

$$Z = \sum_i \tilde{\pi}_i/q_i = \left\{\sum_i q_i\pi_i\right\}^{-1}.$$

(5.39)式可有如下的概率解释：$\pi_i$ 是在状态 $i$ 停留的平均频率，它与嵌入链 $\eta$ 在 $i$ 状态的平均停留次数及平均停留的时间成比例.

对常返的情形，与离散时间的强大数律(§3.4)类似，连续时间参数的马氏链也有强大数律. 下面我们来推导它. 令

$$\tau_0^i(\omega) = 0,$$

$$S_k^i(\omega) = \inf\{t > \tau_k^i(\omega); \xi_t(\omega) \neq i\},$$

$$\tau_{k+1}^i(\omega) = \inf\{t > S_k^i(\omega); \xi_t(\omega) = i\},$$

即 $\tau_k^i(\omega)$ 是 $\xi$ 第 $k$ 次达 $i$ 的时间. 由常返性 $P(\tau_k^i(\omega) < \infty) = 1$. 由强马氏性与§3.4中所作类似的讨论，我们就有 $\tau_1^i(\omega), \tau_2^i(\omega), \cdots$ 是同分布的停时列，而且 $\tau_k^i(\omega)$ 以后的 $\xi$ 与在 $\tau_k^i(\omega)$ 之前发生的事独立. 进而，由每一个 $\tau_k^i(\omega)$ 出发考虑 $\xi$：即考察过程 $\{\xi_{t+\tau_k^i(\omega)}; t \geqslant 0\}$，不难得到它们也都是同分布的. 这样 $\{S_k^i(\omega) - \tau_k^i(\omega); k = 1,2,\cdots\}$ 与 $\{\tau_{k+1}^i(\omega) - \tau_k^i(\omega); k = 1,2,\cdots\}$ 分别都是 i.i.d. 序列. 从而以概率 1 地有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{1}^{n}(S_k^i(\omega)-\tau_k^i(\omega))=E(S_1^i(\omega)-\tau_1^i(\omega))$$

$$=E_i\tau-1/q_i,$$

其中 $\tau(\omega)$ 是 $\xi$ 首次跳跃的时间. 另一方面,还有

$$\lim_{n\to\infty}\frac{\tau_1^i(\omega)}{n}=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(\tau_{k+1}^i(\omega)-\tau_k^i(\omega))=E_i(\tau_1^i(\omega)).$$

与 § 3.4 我们所作过的类似,我们可以得到

$$\lim_{n\to\infty}\frac{1}{t}\int_0^t 1_{\{i\}}(\xi_u)\mathrm{d}u=(E_i(\tau_1^i(\omega)))^{-1};$$

$$E_i\tau=\frac{1}{q_iE_i(\tau_1^i(\omega))}.$$

另一方面,将上式得到的数学期望与(5.39)比较得到

$$E_i\left(\frac{1}{t}\int_0^t 1_{\{i\}}(\xi_u)\mathrm{d}u\right)=\frac{1}{t}\int_0^t p_{ii}(u)\mathrm{d}u\xrightarrow{t\to+\infty}\tilde{\pi}_i=\pi_i/(q_iZ),$$

其中 $$Z=\sum_j\pi_j/q_j.$$

但是左方的极限又应为 $\dfrac{1}{q_iE_i(\tau_1^i(\omega))}$,所以 $E_i(\tau_1^i(\omega))=Z/\pi_i$,即

$$\left(\frac{\pi_i}{Z}\right)E_i(\tau_1^i(\omega))=1. \tag{5.40}$$

综上所述,我们总结出以下定理:

**定理 5.3** 设 $\xi$ 是一个互通常返的 Q-过程,满足对 $\forall i$, $q_i=\sum_{j\neq i}q_{ij}>0$,且 $q_i$ 有界. 则

(1) 对任意状态 $i,j$

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t p_{ij}(s)\mathrm{d}s=\pi_j; \tag{5.41}$$

(2) $\sum_i\pi_ip_{ij}(t)=\pi_j$ ($\forall$ 状态 $j$),

$$\sum_i\pi_i=1; \tag{5.42}$$

(3) $\pi=(\pi_1,\pi_2,\cdots)$ 是如下方程的唯一解.

$$\begin{cases}\sum_i \pi_i q_{ij} = 0 \quad (\forall\ j), \\ \sum_i \pi_i = 1;\end{cases} \tag{5.43}$$

(4) $\pi_i = \bar{\pi}_i/(q_i Z)$，$\tilde{\pi}_i = \dfrac{Z}{E_i \tau_1^i(\omega)}$，而且

$$\sum_i \bar{\pi}_i \bar{P}_{ij} = \bar{\pi}_j, \quad Z = \left(\sum\nolimits_i \frac{1}{E_j \tau_1^j(\omega)}\right)^{-1},$$

即 $\bar{\pi}$ 是 $\xi$ 的嵌入链 $\eta$ 的平稳不变分布，$\tilde{\pi}_i$ 与 $\eta$ 从 $i$ 回到 $i$ 平均所需跳跃时间 $E_i\tau_1^i(\omega)$ 成反比.

(5) 若 $f$ 是状态空间上的有界实函数，则

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f(\xi_u)\mathrm{d}u = \sum_i \pi_i f(i). \tag{5.44}$$

**例 5.4** 在上段的例 5.2 中，当 $\alpha < N\beta$ 时，排队过程的平稳不变分布是 $\pi_i = \alpha_i/Z_1$，其中

$$\alpha_i = (\frac{\alpha}{\beta})^i \frac{1}{(N\wedge i)!N^{(i\vee N)-N}}, \quad Z_1 = \sum\nolimits_i \alpha_i.$$

**证明** 对(5.35)中的 $\boldsymbol{Q}$ 解方程组(5.43)得

$$\pi_1 = \pi_0 \frac{\alpha}{\beta} \equiv \pi_0\alpha_1 \quad (\alpha_0 = 1),$$

$$\pi_2 = \frac{1}{2\beta}(-\pi_0\alpha + \pi_1(\alpha+\beta)) = \frac{\pi_0}{2\beta}\frac{\alpha^2}{\beta} \equiv \pi_0\alpha_2,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\pi_i = \pi_0\left(\frac{\alpha}{\beta}\right)^i \frac{1}{i!} \equiv \pi_0\alpha_i \quad (i \leqslant N),$$

$$\pi_{N+1} = \pi_0\left(\frac{\alpha}{\beta}\right)^{N+1} \frac{1}{N!N},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\pi_j = \pi_0\left(\frac{\alpha}{\beta}\right)^j \frac{1}{N!N^{j-N}} \equiv \pi_0\alpha_j \quad (j > N).$$

因此当 $N\beta > \alpha$ 时有

$$\pi_0 = \left(\sum_{j=0}^{\infty} a_j\right)^{-1} \triangleq Z_1^{-1} \quad (\text{当 } N\beta > \alpha \text{ 时有 } Z_1 < +\infty),$$

$$\pi_i = a_i/Z_1 \quad \left(a_i = \left(\frac{\alpha}{\beta}\right)^i \frac{1}{(i \wedge N)!\,N^{i \vee N - N}} \quad (i = 1,2,\cdots)\right).$$

又由(5.40)与(5.39)我们此时有

$$E_i\tau_i(\omega) = 1/\tilde{\pi}_i = \frac{1}{\pi_i q_i} = Z_1/a_i q_i$$

$$= \left(\frac{\beta}{\alpha}\right)^i (N \wedge i)!\,N^{N-(i \vee N)} Z_1/(\alpha + (N \wedge i)\beta).$$

进而还可得到稳态下(即以 $\underset{\sim}{\pi}$ 为初分布)的平均队长 $L$ 为：

$$L = \sum_{k=1}^{\infty} k\pi_k.$$

特别当 $N=1$ 时有

$$L = \frac{\alpha}{\beta - \alpha}.$$

这是因为 $N=1$ 时，$\pi_i = \dfrac{\left(\frac{\alpha}{\beta}\right)^i}{\sum_{i=0}^{\infty}\left(\frac{\alpha}{\beta}\right)^i} = \dfrac{\left(\frac{\alpha}{\beta}\right)^i}{\left(1 - \frac{\alpha}{\beta}\right)^{-1}}$，

即平稳分布为参数 $1 - \dfrac{\alpha}{\beta}$ 的几何分布. 所以

$$L = \sum_{i=1}^{\infty} i\left(\frac{\alpha}{\beta}\right)^i\left(1 - \frac{\alpha}{\beta}\right) = \sum_{i=1}^{\infty} i\left(\frac{\alpha}{\beta}\right)^{i-1} \frac{\frac{\alpha}{\beta}}{\left(1 - \frac{\alpha}{\beta}\right)^{-1}}$$

$$= \frac{\alpha}{\beta}\left(1 - \frac{\alpha}{\beta}\right)\frac{d}{dz}\left(\frac{1}{1 - z}\right)\bigg|_{z=\frac{\alpha}{\beta}} = \frac{\alpha}{\beta}\left(1 - \frac{\alpha}{\beta}\right)\left(1 - \frac{\alpha}{\beta}\right)^{-2}$$

$$= \frac{\alpha}{\beta}\,\frac{\beta}{\beta - \alpha} = \frac{\alpha}{\beta - \alpha}.$$

进而我们还可以计算以下两种平稳等待时间(指从到达开始到接受服务的时间)：

(1) 在已经"达到"平稳(即充分大)的特定时刻 $t$ 来到平稳排队系统的顾客的平均等待时间.

由于排队系统平稳,对于在系统中的顾客(包括在排队的与正在接受服务的顾客)数 $X_t$ 有:

$$P(\omega: X_t = i) = \pi_i.$$

记时刻 $t$ 来到的顾客等待时间为 $W$,于是

$$P_{(i)}(\omega: W = 0) = 1 \quad (\text{当 } i < N \text{ 时}),$$

而当 $i \geqslant N$ 时,此顾客的等待时间恰为 $t$ 后第 $i - N + 1$ 个顾客离开的时间(因为这之前 $N$ 个"服务员"都一直不停地为那前面的 $i - N + 1$ 个顾客中的 $N$ 个服务,而当第 $i - N + 1$ 个顾客结束服务离开,$N$ 个"服务员"中恰有一个空出来为时刻 $t$ 到来那个顾客服务). 再注意到各顾客与服务员的服务时间相互独立,若记第 $k$ 个顾客离开的时间为 $\sigma_k + t$ (不失一般性,可设 $t = 0$ )则当 $i \geqslant N$ 时

$$\begin{aligned} P_i(\sigma_1 \geqslant s) &= P_i(\omega: N \text{ 个"服务员" 无一在 } s \text{ 前结束服务}) \\ &= \int_s^{\infty} \cdots \int_s^{\infty} e^{-\beta u_1} \beta e^{-\beta u_2} \beta e^{-\beta u_N} \beta \mathrm{d}u_1 \mathrm{d}u_2 \cdots \mathrm{d}u_N \\ &= (e^{-\beta s})^N = e^{-N\beta s}. \end{aligned}$$

可见 $\sigma_1$ 遵从 $N\beta$ 为参数的指数分布. 又若 $i \geqslant N + 1$,在第一个顾客结束服务后,$N$ 个"服务员"各自需要多少时间结束手头服务与这以前的情况独立,而且仍相互独立遵从 $\beta$ 为参数的指数分布,即当 $i \geqslant N + k - 1$ 时,$\sigma_k - \sigma_{k-1}, \sigma_{k-1} - \sigma_{k-2}, \cdots, \sigma_2 - \sigma_1, \sigma_1$ 都相互独立同指数分布;因而 $\sigma_k$ 应遵从 $\Gamma(k, N\beta)$ 分布( 参数为 $k$ 与 $N\beta$ 的 $\Gamma$ 分布 ):

$$\begin{aligned} P(\omega: \sigma_k(\omega) \geqslant s) &= \int_s^{\infty} N\beta \frac{(N\beta u)^{k-1}}{(k-1)!} e^{-N\beta u} \mathrm{d}u \\ &= P(\omega: \sigma_k(\omega) > s) \end{aligned}$$

(参见习题). 最后由于原有 $i$ 人排队加上在时间 $t$ 时新来的顾客 1 人共 $i + 1$ 人,除在被服务的 $N$ 人外,需要等待 $i + 1 - N$ 次服务,

所以我们得到：

$$P(\omega: W(\omega) > s) = \sum_{j \geq N} \pi_j P(\omega: \sigma_{j-N+1}(\omega) > s)$$

$$= \sum_{j \geq N} \pi_j \int_s^{\infty} N\beta \frac{(N\beta u)^{j-N}}{(j-N)!} e^{-N\beta u} du = \sum_{j \geq N} \pi_j \sum_{k=0}^{j-N} e^{-N\beta s} \frac{(N\beta s)^k}{k!}$$

$$= \pi_N e^{-N\beta s} \sum_{j \geq N} \left(\frac{\alpha}{N\beta}\right)^{j-N} \sum_{k=0}^{j-N} \frac{(N\beta s)^k}{k!} = \pi_N e^{-N\beta s} \sum_{k=0}^{\infty} \frac{(N\beta s)^k}{k!} \sum_{j \geq N+k} \left(\frac{\alpha}{N\beta}\right)^{j-N}$$

$$= \pi_N e^{-N\beta s} \sum_{k=0}^{\infty} \frac{(\alpha s)^k}{k!} \sum_{j \geq N+k} \left(\frac{\alpha}{N\beta}\right)^{j-N-k} = \pi_N \frac{1}{1 - \frac{\alpha}{N\beta}} e^{-(N\beta - \alpha)s}.$$

需等待的概率为：$P(W > 0) = \pi_N \left(\frac{1}{1 - \frac{\alpha}{N\beta}}\right)$. 平均等待时间为

$$EW = \int_0^{\infty} P(W(\omega) > s) ds = \int_0^{\infty} \pi_N \frac{1}{1 - \frac{\alpha}{N\beta}} e^{-(N\beta - \alpha)s} ds$$

$$= \pi_N \frac{1}{1 - \frac{\alpha}{N\beta}} \cdot \frac{1}{N\beta - \alpha} = \frac{\pi_N N\beta}{(N\beta - \alpha)^2}.$$

特别当 $N = 1$（即只有一条服务线时，$\pi_1 = \frac{\alpha}{\beta}(1 - \frac{\alpha}{\beta})$，且 $\beta > \alpha$ 时，我们有

$$EW = \frac{\alpha}{\beta} \frac{1}{\beta - \alpha}.$$

(2) 在(1)中求出的平均等待时间假设了顾客恰好在某一特定时刻 $t$ 到来. 这似乎很不合理. 事实上，我们真正希望知道的等待时间应为某一已“达到”平稳（充分大）的时刻 $t$ 开始后第一个顾客到达起这个顾客开始接受服务为止的这段时间. 分析与在(1)中的计算的不同之处是：在现在的计算中应该用 $P(X_\sigma = j + 1)$ 代替(1)中的 $\pi_j$（不同在于：$\sigma$是随机时间，并且多加了此新来顾客 1 人），其中 $\sigma$ 是 $t$ 以后第一个顾客到达的时间（以下不妨设 $t = 0$ ）.

但是

$$P(X_\sigma = j+1) = ?$$

注意到 $\sigma(\omega)$ 必须是 $X(\omega)$ 的跳跃时间：我们考虑嵌入链 $\hat{X}_n(\omega) = X_{\tau_n(\omega)}(\omega)$（其中 $\tau_n(\omega)$ 是 $X(\omega)$ 的第 $n$ 次跳跃时间）. 对于 $j \geqslant N$，在平稳状态下我们有

$$P(\omega: X_\sigma = j+1) = \sum_n P(\omega: \sigma(\omega) = \tau_n(\omega), \hat{X}_n(\omega) = j+1)$$

$$= \sum_{n \geqslant 1} P(\omega: \hat{X}_m(\omega) = \hat{X}_{m-1}(\omega) - 1; 1 \leqslant m \leqslant n-1,$$

$$\hat{X}_n(\omega) = \hat{X}_{n-1}(\omega) + 1 = j+1).$$

由于假定 $t$ 时已达到平稳，所以

$$\hat{\pi}_j \triangleq P(X_\sigma^*(\omega) = j+1)$$

$$= \sum_{i \geqslant j} \pi_i \hat{P}_{i,i-1} \cdots \hat{P}_{j-1,j} \hat{P}_{j,j+1} \text{（这里令 } i = j-n+1\text{）}.$$

注意到 $\hat{P}$ 的可逆平稳分布为 $\tilde{\pi}$，而由(5.39)有

$$\pi_i = \frac{\tilde{\pi}_i}{q_i Z}, \qquad \tilde{\pi}_i \hat{p}_{ik} = \tilde{\pi}_k \hat{p}_{ki}.$$

再注意到对于 $i \geqslant N$ 有 $q_i = q_N$，从而有 $\pi_i \hat{p}_{ik} = \pi_k \hat{p}_{ki}$ 所以上式可改写为：

$$\sum_{i=j}^{\infty} \hat{P}_{i-1,i} \hat{P}_{i-2,i-1} \cdots \hat{P}_{j,j+1} \pi_j \hat{P}_{j,j+1}.$$

注意当 $j \geqslant N$ 时 $\hat{P}_{j,j+1} = \dfrac{\alpha}{\alpha + N\beta}$ 与 $i$ 无关，上式和号下 $i$ 项为

$\pi_j \left(\dfrac{\alpha}{\alpha + N\beta}\right)^{i-j+1}$,

$$\hat{\pi}_j = \pi_j \frac{1}{1 - \dfrac{\alpha}{\alpha + N\beta}} \frac{\alpha}{\alpha + N\beta} = \frac{\alpha}{N\beta} \pi_j.$$

设在(2)的模型下的等待时间为 $\tilde{W}$，则类似于(1)有

$$P(\omega:\tilde{W}(\omega)>s)=\sum_{i=N}^{\infty}\hat{\pi}_i P_{i,i+1}(\omega:\sigma_{i+1\to N}(\omega)\geqslant s)$$

$$=\sum_{i=N}^{\infty}\frac{\alpha}{N\beta}\pi_i\sum_{k=0}^{i-N}e^{-N\beta s}\frac{(N\beta s)^k}{k!}$$

$$=\frac{\alpha}{N\beta}P(\omega:W(\omega)>s)\triangleq rP(W>s),$$

这里 $r=\frac{\alpha}{N\beta}$. 于是在这模型下不需要等待的概率为

$$1-P(\tilde{W}>0)=1-rP(W>0),$$

而且 $E\tilde{W}=rEW=r\frac{\pi_N N\beta}{(N\beta-\alpha)^2}$. 可见 $E\tilde{W}<EW$.

当 $\alpha=N\beta$ 时，$\xi$ 常返，但是 $Z_1=+\infty$，所以此时没有和为有限的不变分布，这时 $\xi$ 是零常返的；而当 $\alpha>N\beta$ 时，$\xi$ 是非常返的. 在这两种情形下排队过程都不会达到平稳.

**例 5.5** 在 0 点反射壁的空间齐次的生灭过程，其概率转移密度为.

$$q_{ij}=\begin{cases}\lambda, & j=i+1, i\geqslant 0;\\ \mu, & j=i-1, i\geqslant 1;\\ -(\lambda+\mu) & i=j, i\geqslant 1;\\ -\lambda & i=j=0.\end{cases}$$

考察 $\xi$ 的嵌入链 $\eta$. 它的转移阵为

$$\tilde{p}_{ij}=\begin{cases}\frac{\lambda}{\lambda+\mu}, & j=i+1, i>0;\\ 1, & j=1, i=0;\\ \frac{\mu}{\lambda+\mu}, & j=i-1, i>0;\\ 0, & \text{其他情形}.\end{cases}$$

由第二章结果可以知到 $\mu\geqslant\lambda$ 时 $\tilde{P}$ 是常返的，特别当 $\mu>\lambda$ 时 $\tilde{P}$ 正

常返，这时 $\bar{P}$ 有唯一的平稳不变分布：

$$\bar{\pi}_i = \left(\frac{\lambda}{\mu}\right)^i \Big/ \sum_j \left(\frac{\lambda}{\mu}\right)^j.$$

由于 $q_i = \lambda + \mu$ 与 $i$ 无关我们就得到：

$$\pi_i = \bar{\pi}_i.$$

## §5.5 对称 Q-过程

### 1. 对称 Q-过程与可逆平稳 Q-过程

正如离散参数马氏链一样，由于对称 Q-过程具有一些非常好的性质，使用起来方便，它是 Q-过程最重要的特例之一.

**定义 5.6** 设 $\xi$ 是一个以 $\mathscr{S}$ 为状态空间，以 $Q = (q_{ij})$ 为转移速率阵的 Q-过程. 称 $\xi$ 为 $Q$ 对称的，如果存在一行不全为 0 的非负实数 $\underset{\sim}{\mu} \triangleq \{\mu_i; i \in \mathscr{S}\}$ 使对 $\forall\, i \neq j \in \mathscr{S}$，有

$$\mu_i q_{ij} = \mu_j q_{ji}, \tag{5.45}$$

这时称 $\underset{\sim}{\mu}$ 为 $Q$ 的对称化数列或配称列（也有作者称为对称化分布）.

我们的第一个问题是 $\xi$ 为 $Q$ 对称与其 $Q$ 对称化数列 $\underset{\sim}{\mu}$ 与 $\xi$ 的转移概率阵有何关系？特别当 $\underset{\sim}{\mu}$ 是一个概率分布的时候，以 $\underset{\sim}{\mu}$ 为初分布是否能得到 $\xi$ 是一个可逆平稳过程？

**定理 5.4** 设 $\xi$ 是有限状态的 Q-过程，其转移概率阵族为 $\{P(t); t \geqslant 0\}$，则 $\xi$ 是 $Q$ 对称的且 $Q$ 以 $\underset{\sim}{\mu}$ 为对称化分布当且仅当 $\xi$ 是以 $\underset{\sim}{\mu}$ 为对称化数列的对称 Q-过程，即对 $\forall\, i \neq j \in \mathscr{S}$，及 $t \geqslant 0$ 有

$$\mu_i p_{ij}(t) = \mu_j p_{ji}(t). \tag{5.46}$$

类似于离散参数的情况，这时 $\underset{\sim}{\mu}$ 可归一化为一个概率分布 $\underset{\sim}{v}$

$=\{v_i; i\in \mathscr{S}\}\{v_i=\mu_i/\sum_i \mu_i\}$ 以 $\underset{\sim}{v}$ 为初分布，$\xi$ 变为一个可逆的平稳过程①，即对 $\forall\ T>0$ 及 $\forall\ t_1,t_2\cdots,t_n\leqslant T$，及 $s\geqslant 0$ 有

$$(\xi_{t_1},\xi_{t_2},\cdots,\xi_{t_n})\overset{d}{=}(\xi_{t_1+s},\xi_{t_2+s},\cdots,\xi_{t_n+s})\qquad(5.47)$$

与

$$(\xi_{t_1},\xi_{t_2},\cdots,\xi_{t_n})\overset{d}{=}(\xi_{T-t_1},\xi_{T-t_2},\cdots,\xi_{T-t_n}).\qquad(5.48)$$

即 $\{\xi_t;t\in[0,T]\}\overset{d}{=}\{\xi_{T-t};t\in[0,T]\}$. 一般地用 $\xi\overset{d}{=}\eta$ 表示 $\xi$ 与 $\eta$ 同分布.

**证明** (1) 若 $\underset{\sim}{\mu}$ 满足(5.46)，即对 $i\neq j$

$$\mu_i p_{ij}(t)=\mu_j p_{ji}(t),$$

于是在等式两边求 $t=0$ 的右导数就得到

$$\mu_i q_{ij}=\mu_j q_{ji}.$$

反之，若(5.45)成立，由于

$$p_{ij}(t)=\sum_{n=0}^{\infty}\frac{t^n}{n!}(Q^n)_{ij},$$

为了证明(5.46)我们只要证明对 $\forall n$ 及

$$M\triangleq\begin{pmatrix}\mu_1 & & & 0\\ & \ddots & & \\ & & \mu_n & \\ 0 & & & \ddots\end{pmatrix},\qquad(5.49)$$

恒有

$$MQ^n=Q^nM.\qquad(5.50)$$

事实上，由(5.45)利用归纳法就立即得到(5.50). 从而(5.46)成立. 进而，若令

$$P(\omega;\xi_0(\omega)=i)=\mu_i\quad(i\in\mathscr{S}),$$

则

---

① 若(5.47)对一切 $t_1,\cdots,t_n$ 与 $\forall\ T,s\geqslant 0$ 成立就称 $\xi$ 是一个平稳过程，意思是 $\{\xi_t;t\geqslant 0\}$ 的任何有限维分布在时间推移下不变. 若(5.48)对一切 $t_1,\cdots,t_n\leqslant T$ 成立，就说明将 $\xi$ 的时间倒过来，$\xi$ 的统计性质不变.

$$P(\omega:\xi_t(\omega)=j)=\sum_i \mu_i p_{ij}(t)=\sum_i \mu_j p_{ji}(t)=\mu_j,$$

即 $\mu$ 是 $\xi_t(\omega)$ 的分布不变. 于是对 $\forall\ t_1\leqslant t_2\leqslant\cdots\leqslant t_n\leqslant T, s\geqslant 0$, 有

$$\begin{aligned}
&P(\omega:\xi_{t_1}(\omega)=i_1,\xi_{t_2}(\omega)=i_2,\cdots,\xi_{t_n}(\omega)=i_n)\\
&\quad=\sum_{i_0}\mu_{i_0}p_{i_0i_1}(t_1)p_{i_1i_2}(t_2-t_1)\cdots p_{i_{n-1}i_n}(t_n-t_{n-1})\\
&\quad=\mu_{i_1}p_{i_1i_2}(t_2-t_1)\cdots p_{i_{n-1}i_n}(t_n-t_{n-1})\\
&\quad=P(\xi_0=i_1)p_{i_1i_2}(t_2-t_1)\cdots p_{i_{n-1}i_n}(t_n-t_{n-1})\\
&\quad=P(\omega:\xi_0(\omega)=i_1,\xi_{t_2-t_1}(\omega)=i_2,\cdots,\xi_{t_n-t_1}(\omega)=i_n).
\end{aligned}$$

而且由(5.46)上式还等于

$$\begin{aligned}
&p_{i_2i_1}(t_2-t_1)\mu_{i_2}p_{i_2i_3}(t_3-t_2)\cdots p_{i_{n-1}i_n}(t_n-t_{n-1})\\
&\quad=p_{i_2i_1}(t_2-t_1)p_{i_3i_2}(t_3-t_2)\mu_{i_3}\cdots p_{i_{n-1}i_n}(t_n-t_{n-1})\\
&\quad=p_{i_2i_1}(t_2-u_1)p_{i_3i_2}(t_3-t_2)\cdots\mu_{i_{n-1}}p_{i_{n-1}i_n}(t_n-t_{n-1})\\
&\quad=p_{i_2i_1}(t_2-t_1)p_{i_3i_2}(t_3-t_2)\cdots p_{i_ni_{n-1}}(t_n-t_{n-1})\mu_{i_n}\\
&\quad=\mu_{i_n}p_{i_ni_{n-1}}(t_n-t_{n-1})\cdots p_{i_2i_1}(t_2-t_1)\\
&\quad=P(\omega:\xi_{T-t_n}=i_n,\xi_{T-t_{n-1}}=i_{n-1},\cdots,\xi_{T-t_2}=i_2,\xi_{T-t_1}=i_1)\\
&\quad=P(\omega:\xi_{T-t_1}=i_1,\xi_{T-t_2}=i_2,\cdots,\xi_{T-t_n}=i_n).
\end{aligned}$$

可见 $\{\xi_t:t\geqslant 0\}$ 是平稳的(满足(5.47)),而且满足(5.48).

对可数状态 Q-过程定理 5.5 并不能一般地成立. 但在一些特殊情形下仍有类似结果:

**定理 5.4′** 设 $\xi$ 是一个可数状态 Q-过程,其 Q-矩阵是互通的(即嵌入链互通),又若有非恒零非负数列 $\mu$ 满足(5.45),且 $Q$ 是保守的 $\left(\text{即 } q_i\triangleq -q_{ii}=\sum_{j\neq i}q_{ij}\right)$ 并有

$$\sum_i \mu_i q_i<+\infty, \tag{5.49}$$

则 $\mu$ 满足(5.46). 又如 $\mu$ 是一个概率分布,则 $\xi$ 正常返,且以 $\mu$ 为

初分布时 $\xi$ 是唯一的一个以 $Q$ 为 Q-矩阵的可逆平稳过程(参见附录).

**2. 对称化数列与可逆平稳分布的计算**

首先,我们要指出:给定一个保守的 Q-矩阵 $\left(\text{即}\ q_i \triangleq -q_{ii} = \sum_{j\neq i} q_{ij}\right)$. 又设 $Q$ 是一个有限时间内不会发生无穷次跳跃的 Q-过程 $\xi$ 的转移密度阵(例如满足定理 5.4 及 5.4′情形),要判别 $\xi$ 是否对称,只要考虑 $\widetilde{P} = (q_{ij}/q_i)$ 是否对称(但两者的对称化数列一般不是一样的). 对后者 Kolmogorov 可逆准则(定理 3.23)及其推论给出了如下很方便的充分必要的判别准则:在互通状态形成的任一个基本圈(不含子圈的圈) $i_1 \to i_2 \to \cdots \to i_s \to i_1$ 上有

$$\prod_{k=1}^{s} \frac{\widetilde{p}_{i_k i_{k+1}}}{\widetilde{p}_{i_{k+1} i_k}} = 1 \quad (i_{s+1} = i_1). \tag{5.50}$$

它又等价于

$$\prod_{k=1}^{s} \frac{q_{i_k i_{k+1}}}{q_{i_{k+1} i_k}} = 1 \quad (i_{s+1} = i_1). \tag{5.50$'$}$$

而且这时

$$\mu_i = \prod_{k=0}^{m} \frac{q_{i_k i_{k+1}}}{q_{i_{k+1} i_k}}. \tag{5.51}$$

其中 $i_0 \to i_1 \to \cdots \to i_m \to i_{m+1} = i$ 是任意取定的一个状态 $i_0$ 到状态 $i$ 的一条 $Q$(或 $\widetilde{P}$) 的通路.

由此可见对于给定 $Q$,要判别它是否有对称化列,是否对称,是否可逆只要先检查(5.50)′,如(5.50)′对一切基本圈成立,则(5.51)就给出了对称化数列. 又如(5.49)成立就知道 $\mu$ 满足(5.46)进而如还有

$$\sum_i \mu_i < +\infty, \tag{5.52}$$

则 $\underset{\sim}{\mu}$ 是一个可逆平稳分布，即以 $\underset{\sim}{\mu}$ 为初分布 $\xi$ 是一个可逆平稳过程.

**例 5.6** $M/M/N$ 服务系统的平稳可逆性与稳态下的输出流.

考虑例 5.2 与例 5.4 中 $\alpha \leqslant N\beta$ 的情况，这时 $X$ 常返. 又由于当 $|i-j|>1$ 时 $q_{ij}=0$，所以这里全部基本圈都是形如 $i \to i+1 \to i$ 的二状态圈；显然

$$\frac{q_{i,i+1}q_{i+1,i}}{q_{i+1,i}q_{i,i+1}} = 1 \quad (\forall\ i).$$

于是(5.51)成立. 令

$$\mu_0 = 1, \mu_1 = \frac{\alpha}{\beta}, \quad \mu_2 = \frac{1}{2!}\left(\frac{\alpha}{\beta}\right)^2, \quad \cdots,$$

$$\mu_N = \frac{1}{N!}\left(\frac{\alpha}{\beta}\right)^N,$$

$$\mu_{N+1} = \mu_N \frac{\alpha}{\beta N} = \frac{1}{N!}\left(\frac{\alpha}{\beta}\right)^{N+1}\frac{1}{N},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\mu_{N+i} = \frac{1}{N!}\left(\frac{\alpha}{\beta}\right)^{N+i}\left(\frac{1}{N}\right)^i,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\mu_n = \frac{1}{N!}\left(\frac{\alpha}{\beta}\right)^n \frac{1}{N^{n-N}} \quad (\text{当 } n > N).$$

当 $\alpha < N\beta$ 时我们有

$$Z = \sum_n \mu_n < +\infty,$$

可见若令

$$\pi_i = \mu_i/Z = \frac{1}{(i \wedge N)!}\left(\frac{\alpha}{\beta}\right)^i \frac{1}{N^{(i \vee N)-N}Z}, \qquad (5.53)$$

则 $\underset{\sim}{\pi} = (\pi_1, \pi_2, \cdots)$ 是 $\xi$ 的可逆平稳分布. 它与在例 5.2 中所得的平稳不变分布一致. 但这里计算过程较例 5.2 中的直接解方程 $\underset{\sim}{\pi}Q$

$=0$，要简单得多.

现在利用倒逆过程 $\{X_{T-t}; t\in[0,T]\}$ 来研究在稳态下(即初分布是平稳不变分布 $\pi$ 的条件下)这个 $M/M/N$ 系统的输出过程：

$\eta \triangleq \{\eta_t; t\geqslant 0\}$，$\eta_t$ 是在 $(0,t]$ 中接受过服务离开的顾客人数.

对时刻 $t$ 在系统中的人数 $X_t$ 我们有 $(t\leqslant T)$

$$X_t(\omega)=\xi_t(\omega)-\eta_t(\omega)+X_0(\omega);\quad P(X_0(\omega)=i)=\pi_i, \tag{5.54}$$

其中 $\xi_t(\omega)$ 是 $(0,t]$ 中顾客到来的人数. 而对 $\hat{X}_t\triangleq X_{T-t}$，有

$$\hat{X}_t(\omega)=\hat{\eta}_t(\omega)-\hat{\xi}_t(\omega)+\hat{X}_0,$$

$$\hat{X}_0=X_T,\quad P(\hat{X}_0=i)=\pi_i \tag{5.55}$$

其中 $\hat{\eta}_t(\omega)$ 与 $\hat{\xi}_t(\omega)$ 分别表示在 $[T-t,T)$ 中离开与来到的顾客人数. 我们还知道，$\xi_t(\omega)$ 与 $\hat{\eta}_t(\omega)$ 分别是 $X_t(\omega)$ 与 $\hat{X}_t(\omega)$ 在 $[0,t)$ 中增加 1 的跳跃次数；而 $\eta_t(\omega)$ 与 $\hat{\xi}_t(\omega)$ 分别是减少 1 的跳跃次数. 由 $X$ 与 $\hat{X}$ 同分布(具有相同的统计特性)并且 $T$ 可以任意取定，我们就知道

$$\xi_t(\omega)\overset{d}{=}\hat{\eta}_t(\omega),\quad \eta_t(\omega)\overset{d}{=}\hat{\xi}_t(\omega)$$

而且 $\{(X_t,\xi_t):t\leqslant T\}$ 与 $\{(\hat{X}_t,\hat{\eta}_t):t\leqslant T\}$ 同分布. 也就是说 $\hat{\eta}$ 也是 Poisson 过程，再注意到

$$\eta_t=\hat{\eta}_T-\hat{\eta}_{T-t},\qquad \eta_0=\hat{\eta}_T-\hat{\eta}_T=0,$$

我们有：

$$P(\eta_{t+s}-\eta_s=i)=P(\hat{\eta}_{T-s}-\hat{\eta}_{T-t-s}=i)$$

$$=P(\xi_{T-s}-\xi_{T-t-s}=i)=e^{-\alpha t}\frac{(\alpha t)^i}{i!}.$$

可见 $\eta$ 也是以 $\alpha$ 为参数的 Poisson 过程. (5.54)与(5.55)说明排队系统 $M/M/N$ 在时刻 $t$ 时在系统中的人数 $X_t$，在取可逆平稳分

布为初分布的稳态条件下，除初始值外，是两个同分布的(但并不彼此独立的)Poisson 过程 $\xi$ 与 $\eta$ 之差，而其倒逆过程 $\hat{X}$ 除初始值外，也是两个 Poisson 过程 $\hat{\eta}$ 与 $\hat{\xi}$ 之差. 直观地，由于 $\alpha < N\beta$ 及系统已处于平稳状态，出入相抵，故参数 $\beta$ 的作用就显不出来了.

**例 5.7**(串联排队) 若一个排队系统中有两道工序，每个顾客到时先到第一个服务工序排队，第一工序服务完毕立即到第二工序排队. 假定等待场地是无限的，每个工序只能同时为一个顾客服务，第 $i$ 道工序服务时间遵从 $\beta_i$ 为参数的指数分布，各工序的服务时间独立，先来先接受服务. 我们需要对顾客的等待时间进行分析，以便合理地设置服务人员与设备.

首先我们分析第一工序的排队过程：

**引理 5.5** 在一个稳态平稳的 $M/M/1$ 排队系统中(即 $\beta > \alpha$)有

(1) 目前系统中的人数 $X_t$ 与过去的诸顾客离去时刻独立.

(2) 一个顾客在系统中的逗留时间与它离开前的离去过程独立.

**证明** (1) 顾客来到过程 $\xi$ 是 Poisson 过程，故它是独立增量的，所以它将来($t$ 以后)到的顾客人数(即 $\{\xi_u - \xi_t; u > t\}$)与 $X_t$ 独立. 再由 $X$ 分布的时间可逆性知道，$\hat{X}_t$ 必须与 $\hat{\eta}_{T-t} - \hat{\eta}_{T-u} = \eta_u - \eta_t; u > t\}$ 的将来值独立，也即 $\hat{X}_{T-t} = X_t$ 与原系统过去离去的顾客人数及离开时刻独立.

(2) 由于先到先服务，在时间 $T_1$ 到来的顾客将在系统停留的时间与 $T_1$ 以后的顾客到来情况独立. 所以在 $T_1$ 到来又在 $S_1$ 离去的顾客，在系统的逗留时间 $S_1 - T_1$ 与 $T_1$ 后顾客到来的情况独立. 将此解释到倒逆过程上，就是说 $T$ 前倒逆过程在时刻 $T - S_1$ 到来的(即原系统在 $S_1$ 离去的)“顾客”将在系统中停留的时间(仍为 $S_1 - T_1$) 与 $T - S_1$ 后逆过程“顾客”的到来(即 $S_1$ 之前顾客的离去)

情况相互独立. 即 $S_1$ 时刻离去的顾客在系统的逗留时间与 $S_1$ 之前顾客的离去时刻独立.

于是我们有以下很不直观的结果

**定理 5.6** 在稳态下的串联排队系统具有如下性质:

(1) 同一顾客在两个工序的等待时间是相互独立的.

(2) 在两个工序各自的排队顾客数是独立的几何分布随机变量,而且

$$
\begin{aligned}
&P(n\text{ 人在第一工序},m\text{ 人在第二工序排队}) \\
&= \left(\frac{\alpha}{\beta_1}\right)^n\left(1-\frac{\alpha}{\beta_1}\right)\left(\frac{\alpha}{\beta_2}\right)^m\left(1-\frac{\alpha}{\beta_2}\right).
\end{aligned}
$$

**证明** (1) 由引理 5.5(2)一个指定顾客在第一工序花费的时间与他离开第一工序前离开第一工序的顾客(即在第二工序排队或离开的顾客)情况独立,但后者决定了此顾客要在第二工序等待的时间分布.

(2) 由引理 5.5(1),顾客在第一工序的排队人数与过去从第一工序离去的各时刻独立,后者正是到达第二工序的各时刻,而第二工序的服务时间又与第一工序的情况无关,因此第一工序排队人数与第二工序排队人数是独立的. 又由于在时刻 $t$ 在第一工序有 $n$ 人排队的概率是(5.53)给出的平稳不变分布:

$$
\begin{aligned}
P(X_t=n)=\pi_n &= \left(\frac{\alpha}{\beta_1}\right)^n\Big/\sum_{n=0}^{\infty}\left(\frac{\alpha}{\beta_1}\right)^n \\
&= \left(\frac{\alpha}{\beta_1}\right)^n\left(\frac{1}{1-\frac{\alpha}{\beta_1}}\right)^{-1} = \left(\frac{\alpha}{\beta_1}\right)^n\left(1-\frac{\alpha}{\beta_1}\right).
\end{aligned}
$$

所以

$$
\begin{aligned}
&P(n\text{ 人在第 1 工序},m\text{ 人在第 2 工序}) \\
&= \left(\frac{\alpha}{\beta_1}\right)^n\left(1-\frac{\alpha}{\beta_1}\right)\left(\frac{\alpha}{\beta_2}\right)^m\left(1-\frac{\alpha}{\beta_2}\right).
\end{aligned}
$$

**例 5.8**(具弹性壁的连续两分支过程的平稳不变分布)

本例讨论连续参数分支过程的一个特例,即 $f_0+f_1+f_2=$

1，但在 0 点设“弹性壁”. 即

$$q_{ij}=\begin{cases} i\alpha, & i\geqslant 1, j=i+1; \\ i\beta, & i\geqslant 1, j=i-1; \\ -i(\alpha+\beta), & i=j\geqslant 1; \\ \gamma, & i=0, j=1; \\ -\gamma, & i=0, j=0; \\ 0, & \text{其他} \end{cases}$$

$(\alpha=f_0, \beta=(1-f_0))$. $\gamma$ 可称为弹性壁系数，$\gamma=0$ 称 0 点为吸取壁，$\gamma=+\infty$ 称 0 点为反射壁. 这种情况下 $q_i=i(\alpha+\beta)$ 是无界的，我们在 § 5.4 中的方法不适用. 但由定理 5.4′ 我们仍可以得到：$\beta>\alpha$ 时这个过程正常返而且有可逆平稳分布. 事实上，与例 5.4 一样可证明 $Q$ 有对称化数列是

$$\alpha_1=1,\quad \alpha_2=\frac{\alpha}{2\beta},\quad \alpha_3=\frac{\alpha}{2\beta}\frac{2\alpha}{3\beta}=\left(\frac{\alpha}{\beta}\right)^2\frac{1}{3},\quad \cdots,$$

$$\alpha_n=\left(\frac{\alpha}{\beta}\right)^{n-1}\frac{1}{n},\quad \alpha_0=\frac{\beta}{\gamma}.$$

当 $\alpha<\beta$ 时，

$$\sum_{n=1}^{\infty}\alpha_n=\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{\alpha}{\beta}\right)^{n-1}<+\infty.$$

这时，$\pi_n=\alpha_n/z\,(n=0,1,\cdots)$，其中 $z=\frac{\beta}{\gamma}+\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{\alpha}{\beta}\right)^{n-1}$.

而 $\underset{\smile}{\pi}=(\pi_0,\pi_1,\cdots,\pi_n,\cdots)$ 是一个概率分布，并且

$$\sum_{n=0}^{\infty}\pi_n q_n=\sum_{n=1}^{\infty}\alpha_n q_n/z+\frac{\beta}{\gamma}\gamma/z$$

$$=\sum_{n=1}^{\infty}\left(\frac{\alpha}{\beta}\right)^{n-1}(\alpha+\beta)\Big/z+\beta/z<+\infty\quad(\beta>\alpha).$$

由定理 5.4′ 可见以 $\underset{\smile}{\pi}$ 为初分布，$Q$ 为 Q-矩阵的 Q-过程 $\xi$ 是可逆的.

### 3. 可逆 Q-过程的性质

除了在前面讲到的可逆 Q-过程，我们可以比较简单地求出其平稳不变分布外，可逆 Q 过程还有一个重要的性质. 下面的定理指出了此性质.

**定理 5.7** 设 $\xi$ 是一个保守的对称 Q 过程，状态空间为 $\mathscr{S}$. 现考虑一个 $\mathscr{S}$ 的子集 $A$，令

$$\tilde{q}_{ij} = \begin{cases} q_{ij}, & i,j \in A, \quad i \neq j; \\ 0, & i \in A, j \notin A, i \neq j; \\ -\sum\limits_{\substack{j \neq i \\ j \in A}} q_{ij}, & i,j \in A, \quad i = j, \end{cases}$$

则 $\tilde{Q} = (\tilde{q}_{ij})$ 就对应一个以 $A$ 为状态空间的保守的对称 Q-过程，记为 $\xi^{(A)}$(我们称 $\xi^{(A)}$ 为 $\xi$(或 $Q$) 在 $A$ 上的限制)，而且当 $\xi$ 可逆时，将 $\xi$ 的可逆平稳分布限制在 $A$ 上，即令

$$\pi_i^{(A)} = \pi_i / z_A \quad \left( 其中\ z_A = \sum_{i \in A} \mu_i < +\infty \right)$$

所得的 $A$ 上的概率分布 $\underset{\sim}{\pi}^{(A)}$ 是 $\xi^{(A)}$ 的可逆平稳分布.

**证明** 由于 $\pi_i q_{ij} = \pi_j q_{ji}, \forall\ i,j \in \mathscr{S}$，显然 $\{\pi_i^{(A)}; i \in A\}$ 仍是 $\tilde{Q}$ 的对称化数列. 又由于

$$z_A = \sum_{i \in A} \pi_i \leqslant \sum_{i \in \mathscr{S}} \pi_i = 1,$$

可见 $\underset{\sim}{\pi}^{(A)} \triangleq \{\pi_i^{(A)}; i \in A\}$ 是一个概率分布，并且由定理 5.3 有

$$\pi_i^{(A)} p_{ij}^{(A)}(t) = \pi_j^{(A)} p_{ji}^{(A)}(t),$$

其中 $P^A = (p_{ij}^{(A)}(t))$ 是 $\xi^{(A)}$ 的转移阵. 于是以 $\underset{\sim}{\pi}^{(A)}$ 为初分布以 $\tilde{Q}$ 为转移速率阵的 Q-过程 $\xi^{(A)}$ 是可逆平稳 Q-过程.

**例 5.9** ($M/M/N$ 排队系统的消失过程) 在一个稳定的 $M/M/N$ 排队系统中若新顾客到来时发现排队人数已超过 $L(\geqslant N)$ 就不再进入此系统等待而离开(消失). 这个带限容量 $(L)$ 的排队过程就可看成是由原 $M/M/N$ 排队系统的排队过程限制在 $A$

$= \{i : i \leqslant L\}$ 上的限止过程. 它是以

$$\pi_i^{(A)} = \mu_i \Big/ \sum_{i \leqslant L} \mu_i \left( \mu_i = \frac{1}{(i \wedge N)!} \left( \frac{\alpha}{\beta} \right)^i \frac{1}{N^{(i \vee N) - N}} \right)$$

为平稳分布的可逆 Q-过程.

### 4. 随机 Ising 模型的 Gibbs 态及其 Glauber 动力学的不变测度

事实上在有限格点的随机 *Ising* 模型中(见 § 5. 3. 4 段), 转移速率阵(Q-矩阵)是:

$$Q = (q_{\xi\eta}), \quad q_{\xi\eta} = \begin{cases} C(x, \xi), & \eta = \xi^x; \\ 0, & \eta \neq \xi, \xi^x (\forall\ x \in \mathcal{N}^d), \end{cases}$$

其中

$$C(x, \xi) = \exp(-\beta(H(\xi^x) - H(\xi))) \tag{5.56}$$

而且 Q 保守, 即

$$q_\xi = \sum_{x \in \mathcal{N}^d} C(x, \xi).$$

容易看出对任意的 $x \in \mathcal{N}^d$, 若令

$$\alpha(\xi) = \exp(-\beta H(\xi)),$$

则
$$\alpha(\xi) q_{\xi\eta} = \alpha(\eta) q_{\eta\xi} \quad (\forall\ \xi \neq \eta).$$

这是因为若 $\eta = \xi^x$ 由(5. 56)上式显然成立, 若 $\eta \neq \xi, \xi^x$ ($\forall\ x \in \mathcal{N}^d$), 则 $q_{\xi\eta} = q_{\eta\xi} = 0$. 可见只要取 $Z = \sum_\xi \alpha(\xi)$, 则

$$\mu(\xi) \triangleq \alpha(\xi)/Z \tag{5.57}$$

是一个可逆的不变平稳分布. 而且以 $\mu(\cdot)$ 为初分布, 相应的 Q-过程是以 $\mu(\cdot)$ 为平稳不变分布的可逆平稳过程.

此外, 在不少问题中(5. 56)中的转移速率也常取为

$$C(x, \xi) = (1 + \exp(-\beta(H(\xi) - H(\xi^x))))^{-1}. \tag{5.58}$$

这时

$$\mu(\xi) q_{\xi\xi^x} = \frac{\exp(-\beta H(\xi))}{Z(1 + \exp(-\beta(H(\xi) - H(\xi^x))))}$$

$$= \frac{\exp(-\beta H(\xi) - \beta(H(\xi^x) - H(\xi)))}{Z(1+\exp(-\beta(H(\xi^x) - H(\xi))))}$$

$$= \frac{1}{Z}\exp(-\beta H(\xi^x))\frac{1}{1+\exp(-\beta(H(\xi^x) - H(\xi)))}$$

$$= \mu(\xi^x)q_{\xi^x\xi}$$

从而对 $\forall\ \xi,\eta$

$$\mu(\xi)q_{\xi\eta} = \mu(\eta)q_{\eta\xi},$$

即以 $\mu(\cdot)$ 为初分布，(5.58)决定的转移速率的马氏链是一个平稳的马氏链，并且它是一个可逆的平稳马氏过程.

从上面的两种不同的转移速率(5.56)与(5.58))就可以看出以 $H(\cdot)$ 为能量函数，决定的转移速率(或说 Q-过程)是不唯一的. 上面两种转移速率的取法是统计力学中最常见的.

另一种统计物理中常见的问题是所谓由能量函数 $H(\cdot)$ 决定平稳测度的问题. 注意到若将 $\mu(\cdot)$ 当作 $\mathscr{S} = \{-1,+1\}^{Z^d}$ 上的概率分布，那么对某一个取定的 $\xi$，$\eta$ 除了 $x$ 位置外全和 $\xi$ 相同的条件下，$\xi$ 与 $\eta$ 相同（即 $\xi(x) = \eta(x)$）的条件概率是：

$$P(\eta(x) = \xi(x) \mid 对一切\ u \neq x, \eta(u) = \xi(u))$$

$$= \frac{\frac{1}{Z}\exp(-\beta H(\xi))}{\frac{1}{Z}\exp(-\beta H(\xi)) + \frac{1}{Z}\exp(-\beta H(\xi^x))}$$

$$= \frac{1}{1+\exp(-\beta(H(\xi^x) - H(\xi)))}. \qquad (5.59)$$

在统计物理中给定能量函数 $H(\cdot)$ 下，将满足关系(5.59)的概率分布称为 $H(\cdot)$ 的 *Gibbs* 态.

对于有限格点(即 $|Z^d| < +\infty$)的情况，关系(5.59)唯一地由 $H(\cdot)$ 决定 $\mu(\cdot)$. 因为这时(5.59)即

$$\frac{\mu(\xi)}{\mu(\xi)+\mu(\xi^x)} = \frac{1}{1+\exp(-\beta(H(\xi^x) - H(\xi)))},$$

于是

$$\mu(\xi)/\mu(\xi') = \exp(-\beta H(\xi))/\exp(-\beta H(\xi')).$$

又由于每次改变一个位置 $x$ 上的 $\xi(x)$ 值就可逐次得到 $\mathscr{S} = \{-1, +1\}^{\Lambda^d}$ 中的全部组态. 于是可见

$$\mu(\xi) = \frac{1}{Z}\exp(-\beta H(\xi)),$$

其中 $Z = \sum_{\xi}\exp(-\beta H(\xi))$ 称为配分函数.

在物理中的两个更为深入的重要问题是:

1) 当 $N \to +\infty$ 时(物理上称为取热力学极限), 上述相互作用能量是否仍然有 Gibbs 分布, 这时的 Gibbs 分布是否唯一? 若有多个 Gibbs 分布, 也就是物理上所谓的有相变, 意思是一个 Gibbs 分布表示一个相.

2) 当 $\beta = +\infty$ 时, 上述系统的各状态 $\eta$ 之间往往不互通了. 如 $\eta$ 满足条件

$$H(\eta) \leqslant H(\eta^x) \quad (\text{对一切 } x \in \Lambda^d) \tag{5.60}$$

且 $\beta = +\infty$ 时, 意味着 $q_{\eta\eta^x} = 0$, 也就说从此 $\eta$ 出发, 过程就永远留在 $\eta$. 事实上, $1 \triangleq \{\eta(x) = 1, \forall\, x \in \Lambda^d\}$ 与 $-1 \triangleq \{\eta(x) = -1, \forall\, x \in \Lambda^d\}$ 就是满足(5.60)的两个状态. 但是在 $\beta < +\infty$ 时上述马氏链是一定互通的, 因而是常返的, 而且从 $-1$ 出发经有限时间一定到 $+1$, 反之亦然. 这就意味着这两种情况马氏链的遍历性有本质的不同. 问题就在于当 $\beta \to +\infty$ 时马氏链在状态之间的转移情况渐近性质怎样? 随着 $\beta$ 变得很大马氏链处于 $-1, +1$ 等这样的状态及跃迁情况有什么特点? 其实, 我们可以看到 $H(+1) < H(-1)$, 这就意味着可逆不变分布(它也是平稳分布)在 $-1$ 比在 $+1$ 的概率之比是无穷小(当 $\beta \to +\infty$), 可见 $+1$ 是比 $-1$ 更有吸引力, 更稳定的组态. $+1$ 是所有状态中最稳定的状态, 而 $-1$ 较之 $+1$ 就不那么稳定了, 物理上将它称为亚稳态. 对亚稳态的研究近来有很多新进展, 由于它与蛋白质的高级结构——折叠性(folding), 晶体与液晶研究, 神经网络的联想记忆等许多科学上有

重要长远意义的问题相联系，引起了许多学者的重视与兴趣.

对 1)及 Glauber 动力学动态粒子系统有兴趣的读者进一步可参见文献：

Liggett, T. M. Interacting Particle Systems, Springer-Verlag, 1985.

Georgii, H. O. Gibbs Measures and Phase Transitions, Walter de Gruyter, 1988.

2)的问题的研究，在数学上还方兴未艾，目前除了统计物理方面有 Penrose 与 Lebowitz 的书可参考外，其他的数学结果尚只在文献中可见. 读者可参考：

Penrose, O and Lebocowitz, J. L., Towards a Regorous MolecularTheory of Metastability in Fluctuation Phenomena, 2-Ed, North Holland, 1978.

Schonmann, R. H An Approach to Characterize metastability and Critical Droplets in stochastic Ising Model, Ann IHP PHY and Critical Vol. 55, No. 2, 591－600.

陈大岳，冯建峰，钱敏平，二维随机 Ising 模型的亚稳态性（Ⅰ），（Ⅱ），《中国科学(A 辑)》，第 27 卷，第 6,7 期，504--513，618--623，1997.

## §5.6 附　录

### 1. 更新流与更新次数（更新过程）

在本段中，我们进一步讨论更新问题（参见 §2.5）. 设 $\{T_n\}$ 为独立同分布的正值随机变量列. 令

$$S_n = T_1 + T_2 + \cdots + T_n, \quad S_0 = 0. \tag{5.61}$$

我们假定设备的第 $n$ 次更新（或大检修）时刻为 $S_n$，或者说第 $n$ 次更新的时间间隔为 $T_n = S_n - S_{n-1}$，我们称 $\{T_n\}$ 为**更新流**，而把

$T_n$ 的分布称为更新流的分布.

服务系统顾客的到来,顾客受到服务后的离开,记数器中记录到粒子的到来,某街口机动车的到来,保险公司的索赔,股票市场的暴涨,某种大灾难的到来等等都可以看成更新流.

分布为指数分布的更新流简称为 **Poisson** 流(或指数流),$\Gamma(k,\lambda)$ 分布的更新流简称为 $k$ 阶 **Erlang** 流,$k=1$ 的 Erlang 流就是 Poisson 流. 在 $(0,t]$ 中的更新的次数记为 $N_t$,称为**更新过程**,显见

$$N_t = \sup\{k; S_k \leqslant t\}, \tag{5.62}$$

它等价于

$$S_k = \inf\{t; N_t = k\}, \tag{5.63}$$

也就是

$$\{N_t \geqslant k\} = \{S_k \leqslant t\}. \tag{5.64}$$

显然 $N_t$ 是依赖于 $t \geqslant 0$ 的随机变量族,即是随机过程.

以设备更新为例,如果某人在时间 $t$ 开始在此设备上工作,则设备已更新了 $N_t$ 次,第 $N_t$ 次的更新时刻为 $S_{N_t}$,所以这个刚更新的设备已用了 $t - S_{N_t}$ 时间,而下一次更新的时刻为 $S_{N_t+1}$,也就是该设备还要用 $S_{N_t+1} - t$ 时间. 因此我们称 $t - S_{N_t}$ 为**年龄**,而称 $S_{N_t+1} - t$ 为**余寿**(**剩余寿命**).

1° 更新流为指数流与更新次数为 Poisson 过程的等价性,年龄与余寿的分布.

**定理 A** 下述三个叙述彼此等价:

(1) $\{T_n\}$ 是指数流.

(2) $\forall\, n$,$(S_1,\cdots,S_n)$ 的联合密度为

$$\lambda^n \mathrm{e}^{-\lambda S_n} \mathbf{1}_{\{0<S_1<\cdots<S_n\}}, \tag{5.65}$$

其中 $\mathbf{1}_A$ 表示 $A$ 的示性函数,即

$$\mathbf{1}_A = \begin{cases} 1, & \text{如果 } A \text{ 满足;} \\ 0, & \text{如果 } A \text{ 不满足.} \end{cases}$$

(3) $\{N_t;t\geqslant 0\}$ 是 Poisson 过程(这就是把指数流称为 Poisson 流的原因).

**证明**　(1)⇒(2).

$$P(S_1\leqslant s_1,\cdots,S_n\leqslant s_n)$$

$$=\underset{\begin{cases}t_1,\cdots,t_n>0,\\ t_1\leqslant s_1,\\ t_1+t_2\leqslant s_2,\\ \cdots\\ t_1+\cdots+t_n\leqslant s_n\end{cases}}{\int\cdots\int}\lambda^n\mathrm{e}^{-\lambda(t_1+\cdots+t_n)}\mathrm{d}t_1\cdots\mathrm{d}t_n=\underset{\begin{cases}0<y_1<\cdots<y_n,\\ y_1\leqslant s_1,\\ \cdots\\ y_n\leqslant s_n\end{cases}}{\int\cdots\int}\lambda^n\mathrm{e}^{-\lambda y_n}\mathrm{d}y_1\cdots\mathrm{d}y_n$$

$$=\underset{\begin{cases}y_1\leqslant s_1,\\ \cdots\\ y_n\leqslant s_n\end{cases}}{\int\cdots\int}\lambda^n\mathrm{e}^{-\lambda y_n}\mathbf{1}_{0<y_1<\cdots<y_n}\mathrm{d}y_1\cdots\mathrm{d}y_n.$$

由此可见(2)成立.

另一方面,$\{T_k\}_{k\leqslant n}$ 与 $\{S_k\}_{k\leqslant n}$ 之间的对应是一一的,所以应该有(2)⇒(1).

(2)⇒(3).

先求 $S_n$ 的密度函数 $g_{S_n}(s_n)$,由(2)用归纳法:

$$g_{S_n}(s_n)=\int\cdots\int\lambda^n\mathrm{e}^{-\lambda s_n}\mathbf{1}_{0<s_1<\cdots<s_{n-1}<s_n}\mathrm{d}s_1\cdots\mathrm{d}s_{n-1}$$

$$=\lambda^n\mathrm{e}^{-\lambda s_n}\mathbf{1}_{0<s_n}\int\cdots\int\mathbf{1}_{0<s_1<\cdots<s_{n-1}<s_n}\mathrm{d}s_1\cdots\mathrm{d}s_{n-1}$$

$$=\lambda^n\mathrm{e}^{-\lambda s_n}\mathbf{1}_{0<s_n}\frac{s_n^{n-1}}{(n-1)!}\ (=\Gamma(n,\lambda)\text{ 分布}).\tag{5.66}$$

其次,由 $\{S_k\}$ 与 $N_t$ 的含义及(2),我们有

$$P(N_s=m,N_{s+t}\geqslant m+k)$$

$$=P(S_m\leqslant s<S_{m+1}<\cdots<S_{m+k}\leqslant s+t)$$

$$=\underset{\begin{cases}s_m\leqslant s,\\ s<s_{m+1}<\cdots<s_{m+k}\leqslant s+t\end{cases}}{\int\cdots\int}\lambda^{m+k}\mathrm{e}^{-\lambda s_{m+k}}\mathbf{1}_{0<s_1<\cdots<s_{m+k}}\mathrm{d}s_1\cdots\mathrm{d}s_{m+k}$$

(令 $s'_{m+l} = s_{m+l} - s$，仿(5.66)的求法，我们有下式)

$$= \lambda^{m+k} \int_{0<s_1<\cdots<s_m<s}\cdots\int ds_1\cdots ds_m \int_{0<s'_{m+1}<\cdots<s'_{m+k}<t}\cdots\int e^{-\lambda(s'_{m+k}+s)} ds'_{m+1}\cdots ds'_{m+k}$$

$$= \lambda^{m+k}\frac{s^m}{m!}e^{-\lambda s}\int_0^t e^{-\lambda s'_{m+k}}\frac{(s'_{m+k})^{k-1}}{(k-1)!}ds'_{m+k}$$

$$= \lambda^m \frac{s^m}{m!}e^{-\lambda s}\int_0^t g_{S_k}(u)du. \tag{5.67}$$

于是

$$P(N_{s+t} - N_s \geqslant k) = \sum_{m=0}^{\infty} P(N_{s+t} \geqslant k + m, N_s = m)$$

$$= \sum_{m=0}^{\infty}\lambda^m \frac{s^m}{m!}e^{-\lambda s}\int_0^t g_{S_k}(u)du = \int_0^t g_{S_k}(u)du. \tag{5.68}$$

从而由(5.66)得到

$$\begin{aligned} P(N_{s+t} - N_s = k) &= P(N_{s+t} - N_s \geqslant k) - P(N_{s+t} - N_s \geqslant k+1) \\ &= \int_0^t (g_{S_k}(u) - g_{S_{k+1}}(u))du \\ &= \int_0^t \left[\frac{\lambda^k u^{k-1}}{(k-1)!}e^{-\lambda u} - \frac{\lambda^{k+1}u^k}{k!}e^{-\lambda u}\right]du \\ &= \left[\frac{\lambda^k u^k}{k!}e^{-\lambda u}\right]_0^t = \frac{(\lambda t)^k}{k!}e^{-\lambda t}. \end{aligned}$$

特别 $P(N_t = k) = \frac{(\lambda t)^k}{k!}e^{-\lambda t}$. 因此由(5.67)及(5.68)

$$P(N_s = m, N_{s+t} - N_s \geqslant k) = P(N_s = m)P(N_{s+t} - N_s \geqslant k).$$

从而

$$P(N_s = m, N_{s+t} - N_s = k) = P(N_s = m)P(N_{s+t} - N_s = k).$$

同样的推理可以对 $0 < t_1 < t_2 < \cdots < t_n$ 有

$$\begin{aligned} &P(N_s = m, N_{s+t_1} - N_s = m_1, \cdots, N_{s+t_n} - N_{s+t_{n-1}} = m_n) \\ &= P(N_s = m)P(N_{s+t_1} - N_s = m_1)\cdots P(N_{s+t_n} - N_{s+t_{n-1}} = m_n) \\ &= \frac{(\lambda s)^m}{m!}e^{-\lambda s}\frac{(\lambda t_1)^{m_1}}{m_1!}e^{-\lambda t_1}\cdots\frac{(\lambda(t_n - t_{n-1}))^{m_n}}{m_n!}e^{-\lambda(t_n - t_{n-1})}. \end{aligned}$$

这就说明了 $\{N_t; t\geq 0\}$ 的独立增量性，从而它是 Poisson 过程.

(3)⇒(2). 对 $0<s_1<\cdots<s_n$，取 $h_1,\cdots,h_n$ 充分小（都小于 $\min\limits_k(s_k-s_{k-1})$，$s_0\triangleq 0$）. 按 $\{S_k\}$ 与 $N_t$ 的含义我们有

$$\begin{aligned}
&P(s_1<S_1\leq s_1+h_1,\cdots,s_n<S_n\leq s_n+h_n)\\
&=P(N_{s_1}=0,N_{s_1+h_1}-N_{s_1}=1,N_{s_2}-N_{s_1+h_1}=0,\cdots,N_{s_n+h_n}-N_{s_n}=1)\\
&=P(N_{s_1}=0)P(N_{h_1}=1)P(N_{s_2-s_1-h_1}=0)\cdots P(N_{h_n}=1)\\
&=e^{-\lambda s_1}\lambda h_1e^{-\lambda h_1}e^{-\lambda(s_2-s_1-h_1)}\cdots\lambda h_ne^{-\lambda h_n}+o(h_1h_2\cdots h_n)\\
&=\lambda^ne^{-\lambda s_n}h_1\cdots h_n+o(h_1\cdots h_n).
\end{aligned}$$

除以 $h_1\cdots h_n$ 后令 $h_1,\cdots,h_n\to 0$ 便得 $(S_1,\cdots,S_n)$ 的密度 $\lambda^ne^{-\lambda s_n}$，把约束条件 $0<s_1<\cdots<s_n$ 加进来，最后就得到(2).

下面我们求指数更新流的年龄及余寿的分布. 其主要思路为：用 $\{N_t\}$ 表达有关的事件.

(1) 年龄 $t-S_{N_t}$ 的分布.

$$P(t-S_{N_t}>x)=0\quad(x\geq t).$$

而当 $x<t$ 时我们有

$$\begin{aligned}
P(t-S_{N_t}>x)&=\sum_{k=0}^{\infty}P(N_t=k,S_k<t-x)\\
&=\sum_{k=0}^{\infty}P(N_{t-x}=k,N_t-N_{t-x}=0)\\
&=\sum_{k=0}^{\infty}\frac{(\lambda(t-x))^k}{k!}e^{-\lambda(t-x)}e^{-\lambda x}\\
&=e^{-\lambda x}.
\end{aligned}$$

于是

$$P(t-S_{N_t}\leq x)=\begin{cases}1-e^{-\lambda x}, & 0\leq x<t;\\ 1, & x\geq t.\end{cases}\tag{5.69}$$

由于这个分布函数在 $t$ 处有跳跃值 $e^{-\lambda t}$，而在其他 $(0,t]$ 上有"密度"函数 $\lambda e^{-\lambda x}$，所以

$$E(t-S_{N_t})=e^{-\lambda t}t+\int_0^t\lambda xe^{-\lambda x}\mathrm{d}x=\frac{1-e^{-\lambda t}}{\lambda}.$$

特别

$$ES_{N_t}=\frac{e^{-\lambda t}-1+\lambda t}{\lambda}. \tag{5.70}$$

② 余寿 $S_{N_t+1}-t$ 的分布.

$$\begin{aligned}P(S_{N_t+1}-t\leqslant x)&=\sum_{k=0}^{\infty}P(S_{N_t+1}-t\leqslant x,N_t=k)\\&=\sum_{k=0}^{\infty}P(N_t=k,N_{t+x}-N_t\geqslant 1)\\&=\sum_{k=0}^{\infty}P(N_t=k)P(N_x\geqslant 1)\\&=1-P(N_x=0)=1-e^{-\lambda x},\end{aligned}$$

这说明 $S_{N_t+1}-t$ ～参数为 $\lambda$ 的指数分布，即

$$S_{N_t+1}-t\overset{d}{=}T_1. \tag{5.71}$$

这里记号“$\overset{d}{=}$”表示等式两边的随机变量为同分布的.

2° 一般更新系统与更新过程及余寿.

对于非指数的更新流 $\{T_n\}$ 对应的更新过程 $\{N_t;t\geqslant 0\}$ 就不再具有 Markov 性，除了极少数外也并不能得到 $N_t$ 的分布的明显表达式. 但是对于

$$m(t)\triangleq EN_t, \tag{5.72}$$

在很多情形下都可以由解一个积分方程(它是一个第二类 Fredholm 方程，叫做**更新方程**)得到.

下面我们设**更新间隔** $T_n$ 具有密度 $f(x)$ .

**定理 B** 设 $T_n$ 有密度 $f(x)$ ，则 $m(t)\triangleq EN_t$ 满足如下的更新方程

$$m(t)=\int_0^t f(s)\mathrm{d}s+\int_0^t m(t-s)f(s)\mathrm{d}s, \tag{5.73}$$

简记成 $m=F+m*f$ .

**证明** 注意 $T_1\overset{d}{=}T_n$ 及(5.61)，我们有

$$m(t)=\sum_k kP(N_t=k)$$

$$= \sum_{k=1}^{\infty} k\int_0^t P(N_t = k \mid T_1 = s) f(s)\mathrm{d}s$$

$$= \sum_{k=1}^{\infty} k\int_0^t P(S_k \leqslant t < S_{k+1} \mid T_1 = s) f(s)\mathrm{d}s$$

$$= \sum_{k=1}^{\infty} k\int_0^t P\left\{ \sum_{i=2}^{k} T_i \leqslant t - s < \sum_{i=2}^{k+1} T_i \,\Big|\, T_1 = s \right\} f(s)\mathrm{d}s$$

$$= \sum_{k=2}^{\infty} k\int_0^t P\left\{ \sum_{i=2}^{k} T_i \leqslant t - s < \sum_{i=2}^{k+1} T_i \right\} f(s)\mathrm{d}s$$

$$= \sum_{k=1}^{\infty} k\int_0^t P(S_{k-1} \leqslant t - s < S_k) f(s)\mathrm{d}s$$

$$= \sum_{k=1}^{\infty} k\int_0^t P(N_{t-s} = k - 1) f(s)\mathrm{d}s$$

$$= \int_0^t (m(t-s) + 1) f(s)\mathrm{d}s.$$

对于余寿的分布，我们有

**定理 C** 记 $F_t^{(e)}(x)$ 为余寿的分布函数

$$F_t^{(e)}(x) \triangleq P(S_{N_t+1} - t \leqslant x), \tag{5.74}$$

则 $g_x(t) \triangleq 1 - F_t^{(e)}(x-)$ 满足更新方程

$$g_x(t) = \int_{t-x}^{\infty} f(s)\mathrm{d}s + \int_0^t g_x(t-s) f(s)\mathrm{d}s, \tag{5.75}$$

其中 $f$ 为更新间隔的分布密度.

**证明** 由

$$g_x(t) = P(S_{N_t+1} - t > x)$$

$$= \int_0^{\infty} P(S_{N_t+1} - t > x \mid T_1 = s) f(s)\mathrm{d}s,$$

仿定理 B 的推导便可得到结论.

通过余寿的分布可以给出更新过程的增量的一个关系，即

**定理 D** 对余寿分布 $F_t^{(e)}(x)$ 有

$$P(N_{t+x}-N_t=k)=\begin{cases}1-F_t^{(s)}(x-), & k=0;\\ \int_0^x P(N_{x-u}=k-1)F_t^{(s)}(\mathrm{d}u), & k\geqslant 1\end{cases} \tag{5.76}$$

(在 $S_{N_t+1}-t$ 有分布密度 $f_t^{(s)}(x)$ 时，$F_t^{(s)}(\mathrm{d}s)=f_t^{(s)}(s)\mathrm{d}s$ ).

**证明** 仿定理C的证明便得.

3° Erlang 流的更新过程.

设 $\{T_n\}$ 为 $k$ 阶 Erlang 流，即更新间隔 $T_n$ 遵从 $\Gamma(k,\lambda)$ 分布，其密度函数为

$$f_k(t)\triangleq \lambda e^{-\lambda t}\frac{(\lambda t)^{k-1}}{(k-1)!}1_{(0,\infty)}(t)\quad (k\geqslant 2). \tag{5.77}$$

Erlang 流作为更新流比指数流(Poisson 流)更为合理，因为后者假定更新间隔 $T_n$ 遵从指数分布，因而有

$$P(T_1>t+s\mid T_1>s)=P(T_1>t).$$

也即 $T_1$ 的分布与它的过去的情形($s$ 前的情形)无关，如果把设备的自然更新(损坏)作为更新间隔，作指数分布的假定说明了设备没有折旧，这是不合理的，虽然我们在某些情况下，还可用这种模型作为粗略近似. 当然，我们还希望采用别的比较合理的流. 其中最简单的就是 Erlang 流，因为它的更新过程的母函数可以明显的写出来.

下面我们将推导出 $k$ 阶 Erlang 流 $\{T_n\}$ 的更新过程 $N_t$ 的母函数 $G_k(t,z)$ ：

$$G_k(t,z)\triangleq Ez^{N_t}=\sum_{n=0}^{\infty}P(N_t=n)z^n$$

$$=\sum_{n=0}^{\infty}P(S_n\leqslant t<S_{n+1})z^n\quad (S_0=0,S_n=T_1+\cdots+T_n)$$

$$=\sum_{n=0}^{\infty}(P(t<S_{n+1})-P(t<S_n))z^n$$

$$=\sum_{n=0}^{\infty}(P(S_n\leqslant t)-P(S_{n+1}\leqslant t))z^n\quad (\text{下面令 } k=n+1)$$

$$= P(S_0 \leqslant t) + z\sum_{n=1}^{\infty} P(S_n \leqslant t) z^{n-1} - \sum_{k=1}^{\infty} P(S_k \leqslant t) z^{k-1}$$

$$= 1 + (z-1)\sum_{n=1}^{\infty} P(S_n \leqslant t) z^{n-1}. \tag{5.78}$$

我们注意到 $S_n \sim \Gamma(kn,\lambda)$，并令 $z = y^k$，则有

$$\sum_{n=1}^{\infty} P(S_n \leqslant t) z^{n-1} = \sum_{n=1}^{\infty}\left[\int_0^t \frac{(\lambda x)^{kn-1}}{(kn-1)!}\lambda e^{-\lambda x}\mathrm{d}x\right] y^{k(n-1)}$$

$$= y^{1-k}\int_0^t \lambda e^{-\lambda x}\sum_{n=1}^{\infty}\frac{(\lambda x y)^{kn-1}}{(kn-1)!}\mathrm{d}x$$

$$= y^{1-k}\int_0^t \lambda e^{-\lambda x}\sum_{m+1\in\{kn,n=1,2,\cdots\}}\frac{(\lambda x y)^m}{m!}\mathrm{d}x.$$

记 $k$ 阶原单位根为 $\varepsilon_k(\triangleq e^{\frac{2\pi i}{k}})$，我们有：

$$\frac{1}{k}\sum_{j=0}^{k-1}(\varepsilon_k)^{jl} = \begin{cases} 1, & \text{当 } l \text{ 是 } k \text{ 的倍数；} \\ 0, & \text{当 } l \text{ 不是 } k \text{ 的倍数.} \end{cases}$$

于是

$$\sum_{m+1\in\{kn,n=1,2,\cdots\}}\frac{u^m}{m!} = \sum_{m=0}^{\infty}\frac{u^m}{m!}\frac{1}{k}\sum_{j=0}^{k-1}(\varepsilon_k)^{j(m+1)} = \frac{1}{k}\sum_{j=0}^{k-1}(\varepsilon_k)^j e^{u(\varepsilon_k)^j}.$$

从而

$$\sum_{n=1}^{\infty} P(S_n \leqslant t) z^{n-1} = y^{1-k}\int_0^t \lambda e^{-\lambda x}\frac{1}{k}\sum_{j=0}^{k-1}(\varepsilon_k)^j e^{\lambda x y(\varepsilon_k)^j}\mathrm{d}x$$

$$= y^{1-k}\frac{1}{k}\sum_{j=0}^{k-1}\frac{(\varepsilon_k)^j}{1-y(\varepsilon_k)^j}\left(1-e^{-\lambda t(1-y(\varepsilon_k)^j)}\right). \tag{5.79}$$

由(5.78)(5.79)我们有(注意 $z = y^k$)：

**定理 E** $k$ 阶 Erlang 流的更新过程的母函数为

$$G_k(t,z) = 1 + (z-1)z^{\frac{1}{k}-1}\frac{1}{k}\sum_{j=0}^{k-1}\frac{(\varepsilon_k)^j}{1-z^{\frac{1}{k}}(\varepsilon_k)^j}\left(1-e^{-\lambda t(1-z^{\frac{1}{k}}(\varepsilon_k)^j)}\right). \tag{5.80}$$

特别当 $k=2$ 时，有

$$G_2(t,z) = e^{-\lambda t}\left[\mathrm{ch}(\lambda t\sqrt{z}) + \frac{1}{\sqrt{z}}\mathrm{sh}(\lambda t\sqrt{z})\right]. \tag{5.81}$$

**证明** 取 $k=2$，用 $\varepsilon_2=-1$ 代入(5.80)经过初等运算化简便得(5.81).

**推论** 对于2阶 Erlang 流的更新过程 $N_t$ 有

$$EN_t=\frac{\lambda t}{2}+\frac{1}{4}(e^{-2\lambda t}-1),$$

$$\mathrm{Var}N_t=\frac{\lambda t}{4}-\frac{\lambda t}{2}e^{-2\lambda t}+\frac{1}{4}e^{-\lambda t}\mathrm{sh}(\lambda t)-\frac{1}{4}e^{-2\lambda t}\mathrm{sh}^2(\lambda t).$$

**证明** 用 $EN_t=\frac{\partial G_2(t,z)}{\partial z}\Big|_{z=1}$ 与

$$\mathrm{Var}N_t=\frac{\partial^2 G_2(t,z)}{\partial z^2}\Big|_{z=1}+EN_t-(EN_t)^2$$

可得.

4° 更新过程的几个重要事实.

记

$$\mu=ET_1,\qquad \sigma^2=\mathrm{Var}T_1.$$

对于(参数为 $\lambda$ 的)指数流的更新过程 $N_t$，我们有

$$\mu=\frac{1}{\lambda},\quad \sigma^2=\frac{1}{\lambda^2}.$$

于是有

(1) $\frac{EN_t}{t}=\lambda=\frac{1}{\mu}(\forall\ t)$，$P\left(\frac{N_t}{t}\xrightarrow{t\to\infty}\frac{1}{\mu}\right)=1$；

(2) $\frac{VarN_t}{t}=\lambda=\frac{\sigma^2}{\mu^3}(\forall\ t)$；

(3) $\frac{N_t-\frac{t}{\mu}}{\sqrt{\frac{\sigma^2}{\mu^3}t}}=\frac{N_t-\lambda t}{\sqrt{\lambda t}}$ 近似 $N(0,1)(t\to\infty)$；

(4) $E(N_{t+s}-N_t)=\lambda s=\frac{s}{\mu}(\forall\ t)$.

对应地，在一般更新流的情形，其更新过程 $N_t$ 不再能精确地满足(1)～(4)，而只能在 $t\to\infty$ 时在渐近意义下成立，即对于一般更新过程 $N_t$ 有

(1)′ 若 $\mu = ET_1 < \infty$，则

$$\frac{EN_t}{t} \to \frac{1}{\mu} \quad (t \to \infty), \quad P\left\{\frac{N_t}{t} \to \frac{1}{\mu}\right\} = 1$$

(这一结论在第二章中已经说明过)；

(2)′ 若 $\sigma^2 = \mathrm{Var}T_1 < \infty$，则

$$\frac{\mathrm{Var}N_t}{t} \to \frac{\sigma^2}{\mu^3};$$

(3)′ 若 $\sigma^2 < \infty$，则

$$\frac{N_t - \frac{1}{\mu}}{\sqrt{\frac{\sigma^2}{\mu^3}t}} \text{ 近似 } N(0,1)(t \to \infty);$$

(4)′ 若 $N_1$ 的分布不集中于整格点的某个平移的集合上面，则

$$E(N_{t+s} - N_t) \to \frac{s}{\mu} \quad (t \to \infty).$$

## 2. 关于排队论的注记

1° 一般概念.

排队理论是比更新过程更为复杂的一种概率模型. 它涉及服务系统、交通运输、通讯系统等广泛的应用领域. 应用的需要与模型的数学处理的繁复之间的矛盾，使许多实际工作者经常加进了他们自己的直观近似. 关于排队论的文献至今已有十万之多，而且每年还以两千的数量增加. 这里我们只对一些最基本的框架作一点注记. 排队理论涉及一个系统，它包含：

——输入过程：一般是一个更新流 $\{T_i\}$，称为"顾客"流. 更广地也可是成批的输入流.

——服务设备(服务或服务员)的数目：1 个、$N$ 个或∞个. 通常它们之间是独立地工作的.

——通常服务设备对于不同顾客的服务时间是独立同分布的

随机变量 $\tilde{T}_i$ ,它们与输入过程是独立的.

—— 服务规则:常见的是先来先服务.又可分三种:一种是等待制(在服务设备全用上时,多余的顾客排队等候,这需要有充分的等候设施);第二种是消失制(顾客见到所有服务设备都在忙就离去,例如电话通讯);第三种是有限制的排队,即顾客看到队伍长度已达到 $L$ 则离去(等候设备有限制).除单个服务外还有成批服务(顾客够一定数才服务).

排队系统研究的几个主要指标为:

——排队长度的平稳不变分布与其平均值,方差(等待制,用来为设计等候室作参考).

——设备在平稳状态下的平均忙期、平均闲期.

——平稳态下的拒绝服务的概率(消失制).

——平稳态下的等待时间的分布与平均值,方差(等候制用以为设置服务线提供参考).

——平稳态下的效率(平均能服务的顾客数).

一个排队系统用下面的记号表示:输入流分布/服务流分布/服务线数目.

—— $M$ :表示指数分布.

—— $E_k$ :表示 $k$ 阶 Erlang 分布.

—— $G$:表示一般分布函数 $G(x)$,在输入流中在 $G$ 后加 $I$ 表示是输入流.

—— $D$:表示流为常值.

所以常见的有 $M/M/1, M/M/N, M/M/\infty$, $GI/M/1$, $M/G/1$,等等.

描述排队系统的一个主要量是

——排队过程 $N_t$ :在时刻 $t$ 在系统中的顾客数(包括正在被服务的顾客与排队的顾客).除 $M/M/.$ 情形外, $N_t$ 一般不是马氏(过程)链,有时就要加入过程的附加信息使它们合起来的向量过程有马氏性.

2° 典例.

(1) $M/M/N$ 等待制(见§5.3.3及例5.1及例5.3);

(2) $M/M/N$ 消失制.

这时 $N_t$ 的转移概率速率阵(Q 矩阵)为有限矩阵:

$$\begin{bmatrix} -\lambda & \lambda & & & & \\ \mu & (\lambda+\mu) & \lambda & & & \\ & 2\mu & -(\lambda+2\mu) & \lambda & & \\ & \ddots & \ddots & \ddots & & \\ & & (N-1)\mu & -(\lambda+(N-1)\mu) & \lambda \\ & & & N\mu & -N\mu \end{bmatrix},$$

这个 $Q$ 有配称列 $\underset{\sim}{\mu}$:

$$\begin{cases} \lambda\mu_0 = \mu\mu_1, \\ \lambda\mu_1 = (2\mu)\mu_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \lambda\mu_{N-1} = (N\mu)\mu_n. \end{cases}$$

所以 $\{N_t:t\geqslant 0\}$ 有可逆不变分布 $\underset{\sim}{\pi}$:

$$\pi_k = \frac{(\lambda/\mu)^k}{k!}\left(\sum_{l=0}^{N}\frac{(\lambda/\mu)^l}{l!}\right)^{-1} \quad \text{(Erlang 公式)}.$$

在平稳状态下,当系统中有 $N$ 个顾客时就拒收新的顾客(新来的顾客看到"占线"就离去),所以顾客被拒绝的概率为

$$\pi_N = \frac{1}{N!}(\lambda/\mu)^N\bigg/\left(\sum_{l=0}^{N}\frac{1}{l!}(\lambda/\mu)^l\right).$$

(3) $M/M/\infty$.

这时可以求得 $N_t$ 的分布的母函数 $G(t,z)$:

$$G(t,z) \triangleq \sum_{k=0}^{\infty}P(N_t=k)z^k \triangleq \sum_{k=0}^{\infty}p_k(t)z^k.$$

设 $t=0$ 时系统中的顾客数 $N_0$ 具有分布

$$P(N_0=k)=a_k,$$

那么

$$G(0,z) = \sum_{k=0}^{\infty} a_k z^k.$$

仿照§5.3.3(那里是 $N$ 个服务员，如今是 $\infty$ 个服务员)我们有

$$\begin{cases} p'_0(t) = -\lambda p_0(t) + \mu p_1(t), \\ p'_k(t) = \lambda p_{k-1}(t) - (\lambda + k\mu) p_k(t) + (k+1)\mu p_{k+1}(t) \quad (k \geqslant 1). \end{cases}$$

由此可以算得

$$\frac{\partial G}{\partial t} = \sum_{k=0}^{\infty} p'_k(t) z^k = (z-1)\left(\lambda G - \mu \frac{\partial G}{\partial z}\right).$$

即

$$\frac{\partial G}{\partial t} + \mu(z-1)\frac{\partial G}{\partial z} - \lambda(z-1)G = 0.$$

令

$$G(t,z) = \mathrm{e}^{\frac{\lambda}{\mu}(z-1)(1-\mathrm{e}^{-\mu t})} H(t,z),$$

经过计算，上面的方程化简为

$$\frac{\partial H}{\partial t} + \mu(z-1)\frac{\partial H}{\partial z} = 0.$$

两边乘以 $\mathrm{e}^{-\mu t}$，上式变为

$$\frac{\partial(H,(z-1)\mathrm{e}^{-\mu t})}{\partial(t,z)} = 0,$$

这里左方是 $H(t,z)$ 与 $(z-1)\mathrm{e}^{-\mu t}$ 关于$(t,z)$的 Jacobian. 这说明 $H(t,z)$ 是 $(z-1)\mathrm{e}^{-\mu t}$ 的某个函数(记为 $h$)，即

$$H(t,z) = h((z-1)\mathrm{e}^{-\mu t}).$$

利用初始值

$$\sum_{k=0}^{\infty} a_k z^k = G(0,z) = H(0,z) = h(z-1).$$

于是

$$h(z) = \sum_{k=0}^{\infty} a_k (z+1)^k.$$

从而

$$G(t,z)=e^{\frac{\lambda}{\mu}(z-1)(1-e^{-\mu t})}\sum_{k=0}^{\infty}a_k(1+(z-1)e^{-\mu t})^k.$$

这里 $G(t,z)$ 依赖于 $G(0,z)$（或它的系数）的选取，我们把与 $G(0,z)=z^i$ 相对应的 $p_k(t)$ 改记为 $p_{ik}(t)$，则由此可得：

$$p_{ij}(t)=e^{-\frac{\lambda}{\mu}(1-e^{-\mu t})}\sum_{k=0}^{i\wedge j}C_i^k\left(\frac{\lambda}{\mu}\right)^{j-k}\frac{e^{-k\mu t}(1-e^{-\mu t})^{i+j-2k}}{(j-k)!}.$$

于是

$$p_{0j}(t)=\frac{1}{j!}\left[\left(\frac{\lambda}{\mu}\right)(1-e^{-\mu t})\right]^j e^{-\frac{\lambda}{\mu}(1-e^{-\mu t})}$$

$\left(\text{即参数}\frac{\lambda}{\mu}(1-e^{-\mu t})\text{的 Poisson 分布}\right)$. 这时仿 §5.3.3 可以求得 $\{N_t:t\geqslant 0\}$ 的转移概率速率阵（Q-矩阵）为

$$Q=\begin{pmatrix}-\lambda & \lambda & & & \\ \mu & -(\lambda+\mu) & \lambda & & \\ & 2\mu & -(\lambda+2\mu) & \lambda & \\ & & \ddots & \ddots & \ddots\end{pmatrix},$$

而且 $Q$ 有对称化数列 $\underset{\sim}{\pi}$：

$$\pi_k=e^{-\frac{\lambda}{\mu}}\left(\frac{\lambda}{\mu}\right)^k\Big/k!,$$

即 $\{N_t:t\geqslant 0\}$ 以参数为 $\frac{\lambda}{\mu}$ 的 Poisson 分布为可逆不变测度，同时有 $\lim\limits_{t\to\infty}p_{ij}(t)=\pi_j$.

以上的方法与结论可以推广到 $\lambda$ 依赖于 $t$ 的情形，只要假定

$$\lambda(t)\leqslant\text{常数 }\lambda_0.$$

这时候有

$$\lim_{t\to\infty}\left|P(N_t=k)-e^{-\Lambda(t)}\frac{\Lambda(t)^k}{k!}\right|\to 0,$$

其中

$$\Lambda(t)=\frac{1}{\mu}\int_0^t\lambda(s)\mathrm{d}s.$$

(4) $M/G/\infty$.

这时服务时间 $\tilde{T}_i$ 为独立同分布的任意分布函数. 只要 $P(\tilde{T} > 0) > 0$,且 $m \triangleq E\tilde{T}_1$ 存在,那么可以由不复杂的推导(参见:A. Ya. 欣钦,《公用事业理论的数学方法》,科学出版社,1958 年中译本,第 72 页),就可得到

$$\lim_{t\to\infty} P(N_t = k) = \mathrm{e}^{-\lambda m}\frac{(\lambda m)^k}{k!},$$

意即当 $t \to \infty$ 时, $N_t$ 渐近于参数为 $\lambda m$ 的 Poisson 分布.

(5) $M/M/1$ 有限制的排队系统.

这种情形发生在排队的队伍长度不允许超过某个固定数 $L$ 的情形,其典型例子出现在设备维修,候机室设计等问题中.

## 3. 关于由转移速率阵寻找转移矩阵 $P(t)$ 的注记

我们考虑保守的 $Q$,因为至今还未见在物理上有实用背景的过程具有非保守的转移速率阵.

给定一个保守的 Q-矩阵 $Q$,即

$$Q = (q_{ij}),\quad 0 \leqslant -q_{ii} < \infty,\quad 0 \leqslant q_{ij} < \infty,$$

$$\sum_j q_{ij} = 0 \quad (\forall\, i,j \in \mathscr{S}),$$

它是否能成为某个马氏过程的转移速率阵呢? 也就是是否存在转移矩阵族 $P(t)$,满足

$$P(t+s) = P(t)P(s),\quad P(0) = I,\quad P'(0) = Q.$$

最直观的想法是考虑 $\mathscr{S}$ 为有限集的情形,此时显然有唯一的可能:

$$P(t) = \mathrm{e}^{Qt}.$$

这个 $P(t)$ 又显然满足后退方程 $P'(t) = QP(t)$ 及前进方程 $P'(t) = P(t)Q$,所以对于 $\mathscr{S}$ 为无限集,我们很自然地看看从后退方程能否得到一个如上的转移矩阵. 而解后退方程也可以仿效有限维情形,先把后退方程写成等价的积分方程

$$P(t) = I + \int_0^t QP(s)\mathrm{d}s.$$

然后再求这个积分方程的迭代解. 但是在 $\mathscr{S}$ 无限时直接迭代会引起收敛性上的困难. 所以要利用方程 $P'(t) = QP(t)$ 的分量形式的等价积分方程形式作迭代. 它的分量形式为($q_i \triangleq -q_{ii} > 0$):

$$p'_{ij}(t) = -q_i p_{ij}(t) + \sum_{k \neq i} q_{ik} p_{kj}(t).$$

用积分因子 $e^{q_i t}$ 可得其等价积分方程形式为

$$p_{ij}(t) = e^{-q_i t}\delta_{ij} + \sum_{k \neq i}\int_0^t e^{-q_i s} q_{ik} p_{kj}(t-s)\mathrm{d}s.$$

对此方程作迭代:

$$\begin{cases} p_{ij}^{(0)}(t) \triangleq 0, \\ \cdots\cdots\cdots\cdots \\ p_{ij}^{(n+1)}(t) \triangleq e^{-q_i t}\delta_{ij} + \sum\limits_{k \neq i}\int_0^t e^{-q_i s} q_{ik} p_{kj}^{(n)}(t-s)\mathrm{d}s. \end{cases}$$

$$P^{(n+1)}(t) \triangleq (p_{ij}^{(n+1)}(t)).$$

可以证明

**定理 F** $P^{(n)}(t) \xrightarrow{n \to \infty}$ 某个 $P(t)$, 且这个 $P(t)$ 满足:

$$P(t) \triangleq (p_{ij}(t)),$$

$$P(t+s) = P(t)P(s),\ P'(0) = Q,\ P'(t) = QP(t) = P(t)Q,$$

$$p_{ij}(t) \geqslant 0,\quad \sum_j p_{ij}(t) \leqslant 1$$

(这个 $P(t)$ 也可以由前进方程的迭代得到).

如果我们在 $\mathscr{S}$ 外补充一个吸收态 $\partial$(称为死态), 并定义

$$\tilde{p}_{ij}(t) \triangleq p_{ij}(t) \quad (i, j \in \mathscr{S}),$$

$$\tilde{p}_{i\partial}(t) \triangleq 1 - \sum_{j \in \mathscr{S}} p_{ij}(t) \quad (i \in \mathscr{S}),$$

$$\tilde{p}_{\partial\partial}(t) \triangleq 1,\quad \tilde{p}_{\partial i}(t) \triangleq 0 \quad (i \in \mathscr{S}),$$

那么

$$\bar{P}(t) = (\bar{p}_{ij}(t))_{i,j\in\mathscr{S}\cup\{\partial\}}$$

是一个转移阵，即它是某个马氏过程 $\bar{\xi} = \{\bar{\xi}_t : t \geqslant 0\}$ 的转移阵，这个 $\bar{\xi}$ 的状态空间是 $\mathscr{S} \cup \{\partial\}$，它的转移速率阵在 $\mathscr{S}$ 上的限制就是 $Q$.

这个 $\bar{P}(t)$（或 $\bar{\xi}$）有一个缺点，就是多了一个吸收态 $\partial$. 因而 $Q$ 只能刻画 $\bar{\xi}$ 首达 $\partial$ 前的行为. 我们把 $\bar{\xi}$ 首达 $\partial$ 的时刻记为 $\zeta$，称为 $\bar{\xi}$ 的死亡时间（或爆炸时间）. $\bar{\xi}_t$ 的轨道在 $t<\zeta$ 前是阶梯式的跳跃轨道，但在 $\zeta$ 附近就会跳跃无穷多次从而"爆炸"死亡. $\bar{P}(t)$（或 $\bar{\xi}$）称为 $Q$ 对应的最小 Q-过程.

在物理世界中出现的 Q-过程绝大多数都是死亡时间 $\zeta \equiv +\infty$ 的. 因此在数学上对于什么样的 $Q$，其 $\bar{P}(t)$ 与 $\bar{\xi}$ 对应的 $\zeta = +\infty$（也即 $\forall\, i, \sum_j p_{ij} = 1$）这个问题特别感兴趣.

如果 $P(\zeta < +\infty) > 0$，在时刻 $\zeta$ 处重新开始一个与 $\{\bar{\xi}_t : t < \infty\}$ 独立同分布的版本，则由强马氏性质可以证明新得的过程的转移矩阵与 $\bar{P}(t)$ 不同，但是它们具有相同的转移矩阵 $Q$，所以说此时对应于 $Q$ 的 Q-过程不是唯一的，也即以 $Q$ 为转移矩阵的转移函数族不止一个. 相反的，当 $P(\zeta = \infty) = 1$ 时（即差不多是 $\zeta \equiv +\infty$ 时），我们由跳跃链的结构知道 $\bar{P}(t)$ 应该由 $Q$ 唯一决定. 因此称不死亡的（即 $\zeta \equiv +\infty$）情形为对应于 $Q$ 有唯一的 Q-过程的情形. 从而前面提出的问题可以复述为：保守的 $Q$ 在什么条件下对应的 Q-过程唯一.

这里叙述的是对一般 $Q$ 易于判别的充分条件. 至于特殊形式的 $Q$，就可给出更细微的条件.

**定理 G** 若 $Q$ 保守，且 $\exists\, C$ 使

$$q_i \leqslant C \quad (\forall i \in \mathscr{S}).$$

则定理 F 中的 $P(t)$ 是转移矩阵［即 $\sum_j p_{ij}(t) = 1(\forall i)$］，因而 Q 过程唯一.

对于 Q 具有配称列 $\mu$ 的情形，定理 5.5′给出了 Q-过程唯一的一个充分条件，其证明可参阅：

钱敏平：《平稳马氏链的可逆性》，《北京大学报》(1978)，第四期，1—9.

**定理 H** 保守的 $Q$，令 $\tilde{P} = (\tilde{p}_{ij})$，

$$\tilde{p}_{ij} = \begin{cases} \delta_{ij}, & q_i = 0; \\ \dfrac{q_{ij}}{q_i}, & q_i > 0, i \neq j; \\ 0, & q_i > 0, i = j, \end{cases}$$

那么 $\tilde{P}$（正或零）常返当且仅当对应于 $Q$ 的最小 Q-过程 $P(t)$（正或零）常返（此时必然唯一）.

还有一种用无穷维线性方程组的解的性质来表达的唯一性条件，在有些情形中也能成为实际上有效的判据.

## 4. 关于生灭过程的注记

具有弹性壁的连续时间两分支过程（例 5.5）及 $M/M/N$ 系统在 $t$ 时刻的人数 $N_t$，都是单侧生灭过程的特例. 这里所谓单侧是指 $\mathscr{S} = \{0,1,2,\cdots\}$ 的情形（相反的，若 $\mathscr{S} = \{\cdots,-1,0,1,2,\cdots\}$，则称为双侧情形）. 而所谓生灭过程是指其转移速率阵 $Q$ 只在主对角线及其两侧的三条斜线上不等于零的 Q-过程. 我们只注意 $Q$ 保守的情形，即若

$$q_{ij} = \begin{cases} \mu_i, & i \geqslant 1, j = i-1; \\ \lambda_i, & i \geqslant 0, j = i+1; \\ -(\lambda_i + \mu_i), & i \geqslant 1, j = i; \\ -\lambda_0, & i = j = 0, \end{cases}$$

则对应的 Q 过程称为(单侧的)生灭过程.

除了两分支过程及 $M/M/N$ 系统外,线性生灭过程也是生灭过程的重要例子.

1° 线性生灭过程(生灭迁移过程).

这种过程的

$$\lambda_i = a + i\lambda, \quad \mu_i = i\mu \quad (a, \lambda, \mu, i \geqslant 0),$$

其中 $\lambda$ 称为生长率,$\mu$ 称为灭绝率,$a$ 称为迁移率. 求这种 Q-过程对应的最小 Q-过程 $P(t)$ (对应于定理 $F$ 中的 $P(t)$ ),可以由定理 $F$ 通过后退方程或前进方程得到. 我们将用前进方程给出 $\{p_{ij}(t); j \in \mathscr{S}\}$ 的母函数

$$P_i(t,z)\left(\triangleq \sum_{j=0}^{\infty} p_{ij}(t) z^j\right).$$

前进方程 $P'(t) = P(t)Q$ 的分量形式为

$$p'_{ij}(t) = \begin{cases} -a p_{i_0}(t) + p_{i_1}(t), & j = 0; \\ -(a + \lambda j + \mu j) p_{ij}(t) + (a + \lambda(j-1)) p_{i,j-1}(t) \\ \qquad + \mu(j+1) p_{j,j+1}(t), & j \geqslant 1. \end{cases}$$

注意 $p_{ij}(0) = \delta_{ij}$ ,于是 $P_i(t,z)$ 满足

$$\begin{cases} \dfrac{\partial P_i}{\partial t} = a(z-1)P_i + (\mu - \lambda z)(1-z)\dfrac{\partial P_i}{\partial z}, \\ P_i(0,z) = z^i. \end{cases}$$

这是一个一阶线性偏微分程,可以用特征线法求解. 先把方程写成

$$\frac{\partial P_i}{\partial z} + \frac{-1}{(\mu - \lambda z)(1-z)} \frac{\partial P_i}{\partial t} = \frac{a}{\mu - \lambda z} P_i. \qquad (\#)$$

然后引入一条(实际上是一族)曲线 $t = t(z)$ ,使 $P_i$ 在 $t(z)$ 上可以解成 $z$ 的函数. 我们想要选 $t(z)$ ,使方程左方成为 $\dfrac{\mathrm{d}P_i(t(z),z)}{\mathrm{d}z}\left(即 \dfrac{\partial P_i}{\partial t} t'(z) + \dfrac{\partial P_i}{\partial z}\right)$. 为此我们只需选 $t(z)$ 使

$$t'(z) = \frac{-1}{(\mu - \lambda z)(1-z)}$$

$$
= \begin{cases} \dfrac{1}{\lambda(1-z)^2}, & \lambda-\mu=0; \\ -\dfrac{1}{(\lambda-\mu)}\left(\dfrac{1}{1-z}-\dfrac{\lambda}{\mu-\lambda z}\right), & \lambda\neq\mu \end{cases}
$$

(注意若 $\lambda=\mu=0$，则 Q 过程为强度为 $a$ 的 Poisson 过程). 解出 $t(z)$：

$$
t(z) = C + \begin{cases} \dfrac{1}{\lambda(z-1)}, & \lambda-\mu=0, \\ -\dfrac{1}{(\lambda-\mu)}\log\left|\dfrac{\mu-\lambda z}{1-z}\right|, & \lambda\neq\mu \end{cases}
$$

$$
\triangleq C + t_0(z).
$$

这样沿曲线族 $t(z)$ 方程(#)变为可分离系数的方程：

$$
\frac{\mathrm{d}P_i}{\mathrm{d}z} = \frac{a}{\mu-\lambda z}P_i.
$$

此方程的解为

$$
P_i = \begin{cases} Kz^{-a/\lambda}, & \lambda>\mu=0; \\ Ke^{\frac{a}{\mu}z}, & \mu>\lambda=0; \\ K|\mu-\lambda z|^{-\frac{a}{\lambda}}, & \lambda+\mu>0 \text{ 的其他情形}. \end{cases}
$$

由于这解是沿依积分常数 $C$ 的曲线族 $t(z)$ 求的，所以这个任意常数 $K$ 必须依赖于 $C$，即应该有

$$
K = K(C) = K(t-t_0(z)).
$$

于是

$$
P_i(t,z) = \begin{cases} K(t-t_0(z))z^{-a/\lambda}, & \lambda>\mu=0; \\ K(t-t_0(z))e^{\frac{a}{\mu}z}, & \mu>\lambda=0; \\ K(t-t_0(z))|\mu-\lambda z|^{-\frac{a}{\lambda}}, & \lambda+\mu>0 \text{ 的其他情形}. \end{cases}
$$

为了确定函数 $K$ 的具体形式，需要用初值 $P_i(0,z)=z^i$，这样

$$
K(-t_0(z)) = \begin{cases} z^{i+a/\lambda}, & \lambda>\mu=0; \\ z^i e^{-\frac{a}{\mu}z}, & \mu>\lambda=0; \\ z^i|\mu-\lambda z|^{a/\lambda}, & \lambda+\mu>0 \text{ 的其他情形}. \end{cases}
$$

作变换

$$u = -t_0(z) \quad (\text{即 } z = t_0^{-1}(-u)),$$

我们得到：

$$K(u) = \begin{cases} (t_0^{-1}(-u))^{i+\frac{a}{\lambda}}, & \lambda > \mu = 0; \\ (t_0^{-1}(-u))^{i} e^{-\frac{a}{\mu} t_0^{-1}(-u)}, & \mu > \lambda = 0; \\ (t_0^{-1}(-u))^{i} |\mu - \lambda t_0^{-1}(u)|^{a/\lambda}, & \lambda + \mu > 0 \text{ 的其他情形}. \end{cases}$$

把它代入 $P_i(t,z)$ 的表达式，最后得明显表达式

$$P_i(t,z) = \begin{cases} (t_0^{-1}(t_0(z)-t))^{i+a/\lambda} z^{-a/\lambda}, \ \lambda > \mu = 0; \\ (t_0^{-1}(t_0(z)-t))^{i} exp(\frac{a}{\mu}[z - t_0^{-1}(t_0(z)-t)]), \ \mu > \lambda = 0; \\ (t_0^{-1}(t_0(z)-t))^{i} \left| \frac{\mu - \lambda t_0^{-1}(t_0(z)-t)}{\mu - \lambda z} \right|^{\frac{a}{\lambda}}, \ \lambda + \mu > 0 \text{ 的其他情形}, \end{cases}$$

其中

$$t_0(z) = \begin{cases} \frac{1}{\lambda(z-1)}, & \lambda = \mu > 0; \\ \frac{1}{\lambda - \mu} \log \left| \frac{\mu - \lambda z}{1 - z} \right|, & \lambda \neq \mu \end{cases}$$

（如果 $t_0(z)$ 在 $z = \mu/\lambda$ 附近分成两支，前面的 $t_0^{-1}(t_0(z)-t)$ 中的 $t_0^{-1}$ 需取 $t_0(z) - t$ 所在的分支）.

如果 $a = 0$（无迁移），那么 $P_0(t,z) = 1$，即

$$p_{0j}(t) = \begin{cases} 1, & j = 0; \\ 0, & j > 0. \end{cases}$$

此时 0 是吸收态，于是 $p_{j0}(t)$ 是 $t$ 的非降函数.

设 $T_0$ 是 $\bar{P}(t)$ 对应的过程 $\bar{\xi}$ 首达 0 的时刻：

$$T_0 = \inf\{t > 0 : \bar{\xi}_t = 0\} \quad (\inf\varnothing \triangleq +\infty),$$

那么物种灭绝的概率即 $P(T_0 < \infty)$. 又因为 0 是吸收态，所以（注意 $a = 0$）

$$P(T_0 \leqslant t \mid \bar{\xi}_0 = i) = p_{i0}(t) = P_i(t,0)$$

$$= (t_0^{-1}(t_0(0) - t))' = \begin{cases} \left[t_0^{-1}\left(-t - \dfrac{1}{\lambda}\right)\right]^i, & \lambda = \mu > 0; \\ \left[t_0^{-1}\left(\dfrac{1}{\lambda - \mu}\log\mu - t\right)\right]^i, & \lambda \neq \mu. \end{cases}$$

从而

$$P(T_0 < \infty \mid \bar{\xi}_0 = i) = (t_0^{-1}(-\infty))^i.$$

注意到

$$t_0(1) = -\infty \quad (\lambda = \mu > 0),$$

$$t_0\left(\frac{\mu}{\lambda}\right) = -\infty \quad (\lambda \neq \mu).$$

于是

$$t_0^{-1}(-\infty) = \begin{cases} 1, & \lambda = \mu > 0; \\ \dfrac{\mu}{\lambda}, & \lambda \neq \mu. \end{cases}$$

因此

$$P(T_0 < \infty \mid \bar{\xi}_0 = i) = \begin{cases} 1, & \lambda = \mu > 0; \\ \left(\dfrac{\mu}{\lambda}\right)^i, & \lambda \neq \mu. \end{cases}$$

此外，经过仔细计算还可得

$$E(T_0 \mid \bar{\xi}_0 = i) = \int_0^\infty P(T_0 > t \mid \bar{\xi}_0 = i)\mathrm{d}t$$

$$= \begin{cases} +\infty, & \lambda \geqslant \mu; \\ -\dfrac{1}{\lambda}\log(1 - \lambda/\mu), & \lambda < \mu \text{ 且 } i = 1 \end{cases}$$

（实际上，若记 $P_i(t,z) = \sum_{j=0}^{\infty} p_{ij}(t) z^j$，则当 $a = 0$ 时，由前面的公式我们可得到

$$P_1(t,z) = \begin{cases} \dfrac{\mu[1 - e^{(\lambda-\mu)t}] - z[\lambda - \mu e^{(\lambda-\mu)t}]}{[\mu - \lambda e^{(\lambda-\mu)t}] - \lambda z[1 - e^{(\lambda-\mu)t}]}, & \lambda \neq \mu; \\ \dfrac{\lambda t + (1 - \lambda t)z}{(1 + \lambda t) - \lambda t z}, & \lambda = \mu \end{cases}$$

及

$$P_i(t,z) = (P_1(t,z))^i.$$

特别当 $a = 0, \mu = 0$（线性纯生过程）时，得

$$p_{ij}(t) = \begin{cases} C_{j-1}^{i-1}e^{-i\lambda t}(1-e^{-\lambda t})^{j-i}, & j \geqslant i \geqslant 1; \\ 1, & i = j = 0; \\ 0, & \text{其他}. \end{cases}$$

再则，不论 $a(\geqslant 0)$ 为多少，在 $\mu > 0$ 时 $Q$ 总有对称化序列 $\underset{\sim}{\mu} = \{\mu_i; i \geqslant 0\}$：

$$\begin{cases} \mu_0 a = \mu_1 \mu, \\ \cdots\cdots\cdots \\ \mu_{n-1}(a+n\lambda) = \mu_n \cdot n\mu, \\ \cdots\cdots\cdots \end{cases}$$

即

$$\begin{cases} \mu_1 = \dfrac{a}{\mu}\mu_0, \\ \cdots\cdots\cdots \\ \mu_n = \dfrac{a+(n-1)\lambda}{n\mu}\mu_{n-1} = \cdots = \dfrac{(a+(n-1)\lambda)(a+(n-2)\lambda)\cdots a}{n!\mu^n}\mu_0, \\ \cdots\cdots\cdots \end{cases}$$

所以当 $\dfrac{\lambda}{\mu} < 1$ 时，定理 5.5′的条件 $\sum\limits_i q_i\mu_i < \infty$ 满足. 由此可知 $P(t)$ 是转移矩阵（从而 $\bar{P}(t) = P(t)$），而且以 $\underset{\sim}{\pi} = \{\pi_i; i \geqslant 0\}$：

$$\pi_n = \frac{(a)_{n\lambda}}{n!\mu^n}\left(\sum_k \frac{(a)_{k\lambda}}{k!\mu^k}\right)^{-1}$$

为可逆不变分布，其中

$$(a)_{n\lambda} = \begin{cases} a(a+\lambda)\cdots(a+(n-1)\lambda), & n \geqslant 1; \\ 1, & n = 0 \end{cases}$$

称为 Pochammer 记号.

当 $\lambda = \mu$ 时有 $t_0(z) = \dfrac{1}{\lambda(z-1)}$，由此 $t_0^{-1}(u) = 1 + \dfrac{1}{\lambda u}$，从

而

$$f_1^{-1}(f_1(z)-t) = 1 + \left[\lambda\left(\frac{1}{\lambda(z-1)} - t\right)\right]^{-1}$$

$$= \frac{1+(1-\lambda t)(z-1)}{1-\lambda t(z-1)}.$$

特别地让 $z=1$，就得到 $f_1^{-1}(f_1(1)-t)=1$，于是

$$P_1(t,1)=1.$$

这说明当 $\lambda=\mu, a\geqslant 0$ 时 $P(t)$ 是转移矩阵(即 Q 过程是唯一的).
此时当 $i\neq j$ 时有

$$p_{ij} \triangleq \frac{q_{ij}}{q_i} = \begin{cases} \dfrac{n\lambda}{a+2n\lambda}, & j=i-1; \\ \dfrac{a+n\lambda}{a+2n\lambda}, & j=i+1. \end{cases}$$

其中右边的数当 $n$ 大时都近似于 $\frac{1}{2}$，直观上可以想象嵌入链 $\bar{P}$ 在 $n$ 大时近似于对称简单随机徘徊，因而是零常返的(实际上也是对的)，这时由定理 G，$P(t)$ 也是零常返的.

类似的理由可以说明当 $\lambda>\mu$ 时，对于 $P(t)$ (即 $\bar{\xi}$)有：除 $\partial$ (参见本书记号)为吸收态外，$\mathscr{S}$ 中的状态均为暂态.

2° (单侧)生灭过程.

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & & \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & \\ & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 \\ & \ddots & \ddots & \ddots \end{pmatrix}, \quad \mathscr{S}=\{0,1,2,\cdots\}.$$

Q 有对称化序列 $\underset{\smile}{\alpha}=\{\alpha_i: i\geqslant 0\}$：

$$\begin{cases} \alpha_0\lambda_0 = \alpha_1\mu_1, \\ \alpha_1\lambda_1 = \alpha_2\mu_2, \\ \cdots\cdots\cdots \end{cases}$$

即

$$
\begin{cases}
\alpha_1 = \dfrac{\lambda_0}{\mu_1}\alpha_0, \\
\alpha_2 = \dfrac{\lambda_0}{\mu_1}\dfrac{\lambda_1}{\mu_2}\alpha_0, \\
\cdots\cdots\cdots \\
\alpha_n = \left(\prod_{k=1}^{n}\dfrac{\lambda_{k-1}}{\mu_k}\right)\alpha_0, \\
\cdots\cdots\cdots
\end{cases}
$$

当

$$
z \triangleq 1 + \sum_{n=1}^{\infty}\left(\prod_{k=1}^{n}\frac{\lambda_{k-1}}{\mu_k}\right) < \infty
$$

时，$\pi \triangleq \{\pi_i; i \geq 0\}$（$\pi_i \triangleq \alpha_i/z$）是 $Q$ 的一个可逆不变分布.

3°（双侧）生灭过程.

$$
q_{ij} = \begin{cases}
-(\lambda_i + \mu_i), & j = i; \\
\mu_i, & j = i-1; \quad i = 0, \pm 1, \pm 2, \cdots. \\
\lambda_i. & j = i+1,
\end{cases}
$$

此时完全类似地可以得到可逆的充分条件.

4° 双侧反射壁的生灭过程.

典型例子出现在设备维修情形. 例如：

有 $M$ 台设备独立地工作，每台设备的正常运转时间段为服从参数 $\lambda$ 的随机变量，运转不正常就需检修，设有 $N(N \leqslant M)$ 个修理工. 记 $N_t$ 为时刻 $t$ 不能正常运转的设备数（正在修理的与待修的）. 修理工独立地以服从参数 $\mu$ 的随机时间修好设备. 求 $N_t$ 的平稳分布.

**解** 由指数流的性质我们有

$$
P\left\{\text{每台设备在}(t, t+h]\text{中需检修次数}\begin{cases}=1\\ \geqslant 2\end{cases}\right\} = \begin{cases}\lambda h + o(h); \\ o(h),\end{cases}
$$

$$
P\left\{\text{每台在修设备在}(t, t+h]\text{中}\begin{cases}\text{被}\\ \text{未}\end{cases}\text{修好}\right\} = \begin{cases}\mu h + o(h); \\ 1 - \mu h + o(h).\end{cases}
$$

再利用设备运转与检修时间之间的独立性就可求得 $N_t$ 是具有下面的转移速率阵的 Q-过程：

$$Q = (q_{ij}),$$

$$q_{ij} = \begin{cases} -M\lambda, & i = j = 0; \\ M\lambda, & i = 0, j = 1; \\ (N \wedge i)\mu, & i > 0, j = i - 1; \\ (M - i)\lambda, & i < N, j = i + 1; \\ N\mu, & i = N, j = N - 1; \\ -N\mu, & i = j = N. \end{cases}$$

这个 $Q$ 不可约且具有配称列 $\underset{\sim}{\pi} = \{\pi_i, 0 \leqslant i \leqslant N\}$，满足

$$\pi_{k+1}(N \wedge k)\mu = \pi_k(M - k)\lambda,$$

即

$$\pi_{k+1} = \frac{(M - k)}{N \wedge k} \frac{\lambda}{\mu}\pi_k = \cdots = \prod_{i=0}^{k} \frac{(M - i)}{(N \wedge i)}\left(\frac{\lambda}{\mu}\right)^{k+1}\pi_0,$$

其中

$$\pi_0 = \left(\sum_{k=0}^{N-1}\left[\left(\frac{\lambda}{\mu}\right)^{k+1}\prod_{i=0}^{k}\frac{M - i}{N \wedge i}\right]\right)^{-1}$$

是所有设备都正常运转的概率．而由此得到

$$\pi_k = \begin{cases} C_M^k\left(\dfrac{\lambda}{\mu}\right)^k\pi_0, & k \leqslant N; \\ C_M^k\dfrac{(N + 1)(N + 2)\cdots k}{N^{k-N}}\pi_0, & N < k \leqslant M. \end{cases}$$

讨论：如果 $N = M$，那么由上面推出

$$\pi_0 = \left(1 + \frac{\lambda}{\mu}\right)^{-M} = \left(\frac{\mu}{\lambda + \mu}\right)^M,$$

$$\pi_k = C_M^k\left(\frac{\lambda}{\mu}\right)^k\left(\frac{\mu}{\lambda + \mu}\right)^M = C_M^k\left(\frac{\lambda}{\lambda + \mu}\right)^k\left(\frac{\mu}{\lambda + \mu}\right)^{M-k},$$

即 $\underset{\sim}{\pi}$ 是 $n = M, p = \dfrac{\lambda}{\lambda + \mu}$ 的二项分布 $B(n, p)$．从而 $\dfrac{\lambda}{\lambda + \mu}$ 可以

解释为在稳定情形下每台设备待修的概率，而 $\frac{\mu}{\lambda+\mu}$ 是每台设备不需检修的概率，称为一台设备的有效度.

如果这 $M$ 台设备组成一个串联系统，那么在稳定情形下系统正常运转的概率即为 $\pi_0$，所以这时(即修理工数目 $N$ = 设备数 $M$ 时)系统的有效度为 $\pi_0 = \left(\frac{\mu}{\lambda+\mu}\right)^M$，是每台设备有效度之积.

又若这 $M$ 台设备组成一并联系统(且有 $M$ 个修理工时)，那么在稳定情形下系统正常运转的概率为 $1-\pi_M$，所以这时系统的有效度为 $1-\pi_M = 1-\left(\frac{\lambda}{\lambda+\mu}\right)^M$.

当然，这 $M$ 台设备还可能接连成其他不同的系统，对应地就可计算在不同系统中的**有效度**——系统正常运转的概率. 而**失效度**则指系统不能正常运转的概率.

在我们这种情形(即 $N=M$)下，不但可以算出 $N_t$ 的平稳分布 $\underset{\sim}{\pi}$，还可以算出 $N_0=0$ 的条件下 $N_t$ 的分布 $\underset{\sim}{p}(t) \triangleq (p_0(t),\cdots,p_M(t))(p_k(t) \triangleq p(N_t=k|N_0=0))$. 事实上，只要利用母函数方法去求前进方程

$$P'(t)=P(t)Q$$

的第一行的解，经过不太繁的计算便可得到 $\underset{\sim}{p}(t)$ 是参数 $n=M, p=\frac{\lambda}{\lambda+\mu}(1-e^{-(\lambda+\mu)t})$ 的二项分布 $B(n,p)$. 从而可以定义一台设备在时刻 $t$ 的有效度为

$$1-\frac{\lambda}{\lambda+\mu}(1-e^{-(\lambda+\mu)t})=\frac{\mu}{\lambda+\mu}+\frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t}.$$

由此也可以计算串联情形、并联情形及其他情形下系统在时刻 $t$ 的有效度及失效度.

如果我们知道了停产的损失费用及设置一个修理工的工资，上面的分析可以帮助我们在稳定情形下设计修理工的最佳配置数目.

## 5. 关于系统有效度的注记

**例 1**（含不同型设备且修理工数与设备数相等的系统）

$n$ 台设备独立地运转. 设第 $k$ 台设备的正常运转时间按参数为 $\lambda_k$ 的指数分布，而当它不正常时需经参数为 $\mu_k$ 的指数分布随机时间修理才能恢复正常运转，设有 $n$ 个修理工，且修理时间与正常运转时间之间独立. 求系统在稳定情形时的有效度.

**解** 先设 $n=2$. 令

$N_t$ = 在时刻 $t$"第一台设备不正常运转的台数"

+ "第二台设备不正常运转的台数" $\times 2$,

于是 $N_t$ 能取 0,1,2,3 四个数，即 $\mathscr{S}=\{0,1,2,3\}$（"0" 表示都正常，"1"、"2" 分别表示第 1,2 台不正常，3 表示都不正常）.

用与前面类似的推导可得 $N_t$ 是具有以下转移速率阵的 Q-过程：

$$Q=\begin{pmatrix} -(\lambda_1+\lambda_2) & \lambda_1 & \lambda_2 & 0 \\ \mu_1 & -(\mu_1+\lambda_2) & 0 & \lambda_2 \\ \mu_2 & 0 & -(\mu_2+\lambda_1) & \lambda_1 \\ 0 & \mu_2 & \mu_1 & -(\mu_1+\mu_2) \end{pmatrix},$$

这个 $Q$ 不可约且具有配称列 $\underset{\sim}{\pi}=(\pi_0,\pi_1,\pi_2,\pi_3)$，满足

$$\pi_0\lambda_1=\pi_1\mu_1,\quad \pi_0\lambda_2=\pi_2\mu_2,\quad \pi_1\lambda_2=\pi_3\mu_2,\quad \pi_2\lambda_1=\pi_3\mu_1.$$

即

$$\pi_1=\frac{\lambda_1}{\mu_1}\pi_0,\quad \pi_2=\frac{\lambda_2}{\mu_2}\pi_0,\quad \pi_3=\frac{\lambda_2}{\mu_2}\pi_1=\frac{\lambda_1}{\mu_1}\frac{\lambda_2}{\mu_2}\pi_0.$$

于是 $N_t$ 的平稳分布 $\underset{\sim}{\pi}$ 应满足

$$\pi_0=\left(1+\frac{\lambda_1}{\mu_1}+\frac{\lambda_2}{\mu_2}+\frac{\lambda_1}{\mu_1}\frac{\lambda_2}{\mu_2}\right)^{-1}=\frac{\mu_1}{\lambda_1+\mu_1}\cdot\frac{\mu_2}{\lambda_2+\mu_2}.$$

可见稳定后的两台设备串接系统的有效度为

$$\pi_0=\frac{\mu_1}{\lambda_1+\mu_1}\cdot\frac{\mu_2}{\lambda_2+\mu_2},$$

它仍为每台设备有效度之乘积（这是由两台设备正常运转时间的

独立性所决定的)，而稳定后的两台设备并联系统的有效度为

$$1-\pi_3=1-\frac{\lambda_1}{\lambda_1+\mu_1}\,\frac{\lambda_2}{\lambda_2+\mu_2},$$

这也是与直观运算完全一致的.

对一般 $n$ 的情形相应结论的正确性也是显见的.

**例 2**(有一个设备不正常就停工且只有一个修理工的系统)

由 $n$ 个设备串联的系统，设各设备的正常工作时间是独立的，第 $k$ 台设备的正常工作时间遵从参数为 $\lambda_k$ 的指数分布. 系统只有一个修理工，当系统有一个设备不正常时就全系统停工并修理该设备(与例 1 有两点不同，在例 1 中全系统并不停工，其他设备还在工作；有 $n$ 个修理工)，第 $k$ 台设备需参数 $\mu_k$ 的指数分布的随机时间修复，修复后系统立即重新工作，设修理时间与正常运转时间之间为独立，求系统在稳定时的有效度.

**解** 令

$$0=\text{系统正常工作},$$
$$k=\text{因第 } k \text{ 台设备在修理而停工}\ (0<k\leqslant n),$$
$$N_t=\text{时刻 } t \text{ 系统所处的状态}.$$

于是 $N_t$ 取值于 $\mathscr{S}=\{0,1,\cdots,n\}$. 仿前可以证明 $N_t$ 是具有下述转移速率阵的 $Q$-过程：

$$Q=\begin{pmatrix} -(\lambda_1+\cdots+\lambda_n) & \lambda_1 & \cdots & \lambda_n \\ \mu_1 & -\mu_1 & & \\ \vdots & & \ddots & \\ \mu_n & & & \mu_n \end{pmatrix},$$

这个 $Q$ 不可约且具有配称列 $\underset{\smile}{\pi}=(\pi_0,\pi_1,\cdots,\pi_n)$，满足

$$\lambda_1\pi_0=\mu_1\pi_1,\quad \cdots,\quad \lambda_n\pi_0=\mu_n\pi_n.$$

所以稳定后的系统的有效度为：

$$\pi_0=\left(1+\frac{\lambda_1}{\mu_1}+\cdots+\frac{\lambda_n}{\mu_n}\right)^{-1}.$$

显见由

$$\left(1+\frac{\lambda_1}{\mu_1}+\cdots+\frac{\lambda_n}{\mu_n}\right)^{-1} \geqslant \left(1+\frac{\lambda_1}{\mu_1}\right)^{-1}\cdots\left(1+\frac{\lambda_n}{\mu_n}\right)^{-1}$$

推得例 2 的有效度确比例 1 的有效度高. 可见在串联情形只需一个修理工,而且停工检修是更为合理的.

**例 3**(只有一个修理工的并联系统)

这时假设设备为 $n$ 个同类型的,即正常运转时间分布都为 $\lambda$ 的指数分布,而修理所需时间分布都为参数 $\mu$ 的指数分布,同样假定它们独立且每台设备独立地工作,所有设备并联,只有一个修理工,求系统在稳定情形下的有效度.

(注:如果设备为不同类型而只有一个修理工,那么系统不再能用 Q 过程的办法来分析,因为此时修理情况依赖于"历史"(先坏的先修)).

**解** 令 $N_t$ 为时刻 $t$ 不在正常运转的设备数,类似前面的方法可以知道 $N_t$ 是有下述转移速率阵的 $Q$-过程:

$$Q=\begin{pmatrix} -n\lambda & n\lambda & & & \\ \mu & -(\mu+(n-1)\lambda) & (n-1)\lambda & & \\ & \ddots & \ddots & \ddots & \\ & \mu & -(\mu+\lambda) & \lambda \\ & & \mu & -\mu \end{pmatrix},$$

这个 $Q$ 不可约且具有配称列 $\underset{\smile}{\pi}=(\pi_0,\pi_1,\cdots,\pi_n)$,满足

$$\mu\pi_1=n\lambda\pi_0,\quad \mu\pi_2=(n-1)\lambda\pi_1,\quad \cdots,\quad \mu\pi_n=\lambda\pi_{n-1}.$$

所以

$$\pi_1=n\left(\frac{\lambda}{\mu}\right)\pi_0,\ \pi_2=n(n-1)\left(\frac{\lambda}{\mu}\right)^2\pi_0,\ \cdots,\ \pi_n=n!\left(\frac{\lambda}{\mu}\right)^n\pi_0.$$

从而 $N_t$ 的平稳分布 $\underset{\smile}{\pi}$ 应满足

$$\pi_0=\left[1+n\left(\frac{\lambda}{\mu}\right)+n(n-1)\left(\frac{\lambda}{\mu}\right)^2+\cdots+n!\left(\frac{\lambda}{\mu}\right)^n\right]^{-1}.$$

系统的有效度为

$$1-\pi_n=1-\left(\frac{\lambda}{\mu}\right)^n\Big/\left[\sum_{k=0}^{n}\frac{1}{(n-k)!}\left(\frac{\lambda}{\mu}\right)^k\right].$$

**例 4**(具有备用设备的系统)

冷贮系统有 $n$ 台相同设备，其中一台在工作，其他 $(n-1)$ 台为备用. 设有一条修理线，设备正常运转时间服从参数为 $\lambda$ 的指数分布，与它独立的修理设备的时间服从参数为 $\mu$ 的指数分布，求系统在稳定时的有效度.

**解** 令 $N_t$ 为时刻 $t$ 在修与待修的设备台数之和，仿前面的推导可得 $N_t$ 是具有下述转移速率阵的 $Q$-过程：

$$Q=\begin{pmatrix} -\lambda & \lambda & & & \\ \mu & -(\lambda+\mu) & \lambda & & \\ & \ddots & \ddots & \ddots & \\ & \mu & -(\lambda+\mu) & \lambda & \\ & & \mu & -\mu \end{pmatrix}(\mathscr{S}=\{0,1,\cdots,n\}).$$

$Q$ 不可约且具有配称列 $\underset{\smile}{\pi}=(\pi_0,\pi_1,\cdots,\pi_n)$，满足

$$\pi_0\lambda=\pi_1\mu,\cdots,\pi_k\lambda=\pi_{k+1}\mu\quad(0\leqslant k<n).$$

所以 $N_t$ 的平稳分布 $\underset{\smile}{\pi}$ 应满足

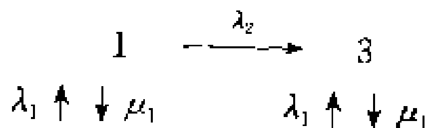
$$\pi_0=\left[1+\frac{\lambda}{\mu}+\cdots+\left(\frac{\lambda}{\mu}\right)^n\right]^{-1}.$$

$\pi_0$ 即是系统在稳定时的有效度.

**例 5**（具有性能较差的备用设备的系统）

冷贮系统由两台设备组成，其中一台是工作设备，另一台是比前一台性能较差的备用设备. 有一个修理工，当失效的工作设备被修好时立刻用它来代替正在工作的备用设备，而若工作设备失效且备用设备正在修理时则立刻停修后者并先修前者. 设工作设备与备用设备的正常运转时间分别服从参数为 $\lambda_1,\lambda_2$ 的指数分布，与它独立的修理设备的时间分别服从参数为 $\mu_1,\mu_2$ 的指数分布，求系统在稳定时的有效度.

**解** 我们定义 $N_t$ 如同例 1，根据题意中优先修理优先使用工作设备的要求，我们画一张示意"流向图"：



$$0 \xleftarrow{\mu_2} 2$$

类似的推导可得到 $N_t$ 是具有下述转移速率阵的 Q 过程：

$$Q = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & 0 \\ \mu_1 & -(\mu_1+\lambda_2) & 0 & \lambda_2 \\ \mu_2 & 0 & -(\mu_2+\lambda_1) & \lambda_1 \\ 0 & 0 & \mu_1 & -\mu_1 \end{pmatrix},$$

这个 $Q$ 不可约但没有配称列，我们可以用 $\underset{\sim}{\pi}Q = 0$ 求 $N_t$ 的平稳分布 $\underset{\sim}{\pi} = (\pi_0, \pi_1, \pi_2, \pi_3)$，也就是：

$$\begin{cases} -\lambda_1\pi_0 + \mu_1\pi_1 + \mu_2\pi_2 = 0, \\ \lambda_1\pi_0 - (\mu_1+\lambda_2)\pi_1 = 0, \\ -(\mu_2+\lambda_1)\pi_2 + \mu_1\pi_3 = 0, \\ \pi_0 + \pi_1 + \pi_2 + \pi_3 = 1, \pi_i \geqslant 0. \end{cases}$$

解此方程组得到：

$$\pi_1 = \frac{\lambda_1}{\lambda_2+\mu_1}\pi_0,$$

$$\pi_2 = \frac{\lambda_1}{\mu_2}\pi_0 - \frac{\mu_1}{\mu_2}\pi_1 = \frac{1}{\mu_2}\left(\lambda_1 - \mu_1\frac{\lambda_1}{\lambda_2+\mu_1}\right)\pi_0 = \frac{\lambda_1}{\mu_2}\cdot\frac{\lambda_2}{\lambda_2+\mu_1}\pi_0,$$

$$\pi_3 = \frac{\lambda_1+\mu_2}{\mu_1}\pi_2 = \frac{\lambda_1\lambda_2}{\mu_2(\lambda_2+\mu_1)}\cdot\frac{\lambda_1+\mu_2}{\mu_1}\pi_0.$$

于是系统在稳定时的有效度为

$$\pi_0 = \left(1 + \frac{\lambda_1}{\lambda_2+\mu_1} + \frac{\lambda_1}{\lambda_2+\mu_1}\cdot\frac{\lambda_2}{\mu_2} + \frac{\lambda_1}{\lambda_2+\mu_1}\cdot\frac{\lambda_2}{\mu_2}\cdot\frac{\lambda_1+\mu_2}{\mu_1}\right)^{-1}$$

$$= \frac{\mu_1\mu_2(\lambda_2+\mu_1)}{\lambda_1\lambda_2(\lambda_1+\mu_1+\mu_2) + \mu_1\mu_2(\mu_1+\lambda_1+\lambda_2)}.$$

**例 6**（系统有处于其他状态的同型备用设备）

热贮系统由两台相同的设备，其中一台处于工作态，另一台处于热贮态. 它们的正常运转时间分别服从参数为 $\lambda, \nu$ 的指数分布，两台设备在不正常时所需修理时间均为参数为 $\mu$ 的指数分布. 有一个修理工，先失效者先修，但是如果工作态设备失效而热贮设备

正常，则后者立刻转为工作态，而且正常运转时间的指数分布的参数由 $v$ 转变为 $\lambda$，此后如果原来处工作态的设备修复，则此修复的设备转为热贮状态，并且其运转时间的指数分布的参数由 $\lambda$ 转变为 $v$，求这个系统在稳定时的有效度.

**解** 令

$$0 = 两台都正常,$$
$$1 = 一台处于工作态,另一台在修理,$$
$$2 = 两台都不正常(系统失效).$$

$\mathscr{S} = \{1,2,3\}$，$N_t$ 为系统在时刻 $t$ 所处的状态，用类似的思路可以推出 $N_t$ 是具有下述转移速率阵的 Q-过程：

$$Q = \begin{pmatrix} -(\lambda+v) & \lambda+\mu & 0 \\ \mu & -(\lambda+v) & \lambda \\ 0 & \mu & -\mu \end{pmatrix},$$

这个 $Q$ 不可约且具有配称列：

$$\underset{\sim}{\pi} = (\pi_0, \pi_1, \pi_2),$$

$$\pi_0(\lambda+v) = \pi_1\mu, \pi_1\lambda = \pi_2\mu.$$

所以 $N_t$ 的平稳分布为 $\underset{\sim}{\pi}$，其中

$$\pi_1 = \frac{\lambda+v}{\mu}\pi_0, \pi_2 = \frac{\lambda(\lambda+v)}{\mu^2}\pi_0,$$

$$\pi_0 = \left(1 + \frac{\lambda+v}{\mu} + \frac{\lambda(\lambda+v)}{\mu^2}\right)^{-1}.$$

于是在系统稳定时的有效度为：

$$1 - \pi_2 = 1 - \frac{\lambda(\lambda+v)}{\mu^2 + (\lambda+v)\mu + \lambda(\lambda+v)}.$$

## 习 题

1. 若随机过程 $N_t$ 取非负整数值，满足

$(P_1)$ $N_0 = 0$;

$(P_2)$ $N_t$ 为独立增量过程，即 $\forall\ 0 < t_1 < \cdots < t_n$ 有 $N_{t_1}, N_{t_2} - N_{t_1}, \cdots, N_{t_n} - N_{t_{n-1}}$ 相互独立；

$(P_3)$ 当 $h \to 0$ 时有

$$\begin{cases} P(N_{t+h} - N_t = 1) = \lambda h + o(h), \\ P(N_{t+h} - N_t \geqslant 2) = o(h). \end{cases}$$

记 $N_t$ 的母函数为

$$G(t,z) \triangleq \sum_{k=0}^{\infty} P_k(t) z^k,$$

其中

$$P_k(t) = P(N_t = k).$$

求证它满足

$$\begin{cases} \dfrac{\partial G}{\partial t} = \lambda(z-1)G, \\ G(0,z) = 1. \end{cases}$$

并由此证明 $G(t,z) = e^{\lambda t(z-1)}$ 及 $N_t$ 是 Poisson 过程.

2. 若随机过程 $X_t$ 取整数值，满足

$(P_1)$ $X_0 = 0$；

$(P_2)$ $X_t$ 独立增量；

$(P_3)$ 当 $h \to 0$ 时有

$$\begin{cases} P(X_{t+h} - X_t = -1) = \mu h + o(h), \\ P(X_{t+h} - X_t = 1) = \lambda h + o(h), \\ P(X_{t+h} - X_t \overline{\in} [-1,1]) = o(h), \end{cases}$$

记 $X_t$ 的母函数为

$$G(t,z) \triangleq \sum_{k=-\infty}^{\infty} P(X_t = k) z^k.$$

求 $G$ 满足的方程，由此证明：

$$G(t,z) = e^{t[-(\lambda+\mu)+\lambda z+\mu\frac{1}{z}]},$$

并且有 $X_{t+s} - X_s \overset{d}{=} X_t$. 再求 $EX_t$, $\mathrm{Var}X_t$.

3. 在习题 1 中的 $(P_3)$ 若改为

$(P'_3)$ 当 $h \to 0$ 时有

$$\begin{cases} P(N_{t+h} - N_t = 1) = \lambda_1 h + o(h), \\ P(N_{t+h} - N_t = 2) = \lambda_2 h + o(h), \\ P(N_{t+h} - N_t \geqslant 3) = o(h). \end{cases}$$

试解出 $N_t$ 的母函数，你能否再进一步推广至

$$\begin{cases} P(X_{t+h} - X_t = i) = \lambda_i h + o(h) \quad (i \leqslant k), \\ P(X_{t+h} - X_t = -j) = \mu_j h + o(h) \quad (j \leqslant l), \\ P(X_{t+h} - X_t \overline{\in} [-l,k]) = o(h) \end{cases}$$

这种情形.

4. 在习题 1 的 $(P_3)$ 中将 $\lambda$ 改为 $\lambda(t)$，求 $G(t,z)$，$EN_t$，$\mathrm{Var}N_t$（对应的过程称为非齐次的 Poisson 过程）.

5. 在习题 2 中的 $(P_3)$ 将 $\lambda,\mu$ 改为 $\lambda(t),\mu(t)$，求 $G(t,z)$.

6. 推广习题 3 中至 $\lambda_i,\mu_i$ 依赖于 $t$ 的情形.

7. 在习题 1 中的 $(P_3)$ 如果改为

$$\begin{cases} P(N_{t+h} - N_t = 1 \mid N_t = k) = \lambda_k h + o(h), \\ P(N_{t+h} - N_t \geqslant 2 \mid N_t = k) = o(h). \end{cases}$$

求 $P_k(t) \triangleq P(N_t = k)$ 满足的微分方程，并解出 $P_0(t)$，$P_1(t)$，$P_2(t)$.

8. 设非负整值的随机过程 $\{N_t; t \geqslant 0\}$ 如满足：

(1) $N_t$ 独立增量，且 $N_{s+t} - N_s \overset{d}{=} N_t$；

(2) $\forall\, t, 0 < P(N_t = 0) < 1$.

求证存在 $0 < \lambda < \infty$，使

$$P(N_t = 0) = e^{-\lambda t}.$$

9. 设非负整值的随机过程 $\{N_t; t \geqslant 0\}$ 如满足习题 8 的条件，而且还满足

$$P(N_h = k \mid N_h > 0) = p_k h + o(h).$$

求证 $N_t$ 的母函数有表达式

$$G(t,z) = e^{\lambda t(g(z)-1)},$$

其中

$$g(z) = \sum_{k=1}^{\infty} p_k z^k.$$

这个 $\{N_t; t \geqslant 0\}$ 也称为广义 Poisson 过程.

提示：令 $P_k(t) = P(N_t = k)$，先建立关系

$$P_k(t+h) = e^{-\lambda h} P_k(t) + (1 - e^{-\lambda h}) \sum_{j=1}^{k} p_j P_{k-j}(t) + o(h).$$

10. 设 $\{T_n\}$ 为 i.i.d.，其分布为参数 $p$ 的几何分布：

$$P(T_n = l) = pq^{l-1} \quad (q = 1 - p, l \geqslant 1).$$

$S_n = T_1 + \cdots + T_n, N_n = \max\{k; S_k \leqslant n\}$，求证

(1) $S_n$ 遵从负二项分布：

$$P(S_n = l) = C_{l-1}^{n-1} p^n q^{l-n} \quad (q = 1 - p, l \geqslant n).$$

(2) $N_n$ 遵从参数为 $p$ 及 $n$ 的二项分布.

(3) $P(S_1 = l_1, \cdots, S_n = l_n) = p^n q^{l_n - n} \mathbf{1}_{l_1 < l_2 < \cdots < l_n}$（再由它得出 $P(S_n = l)$）.

(4) $P(N_n - N_{n-1} \geqslant 2) = 0$，$N_n - N_{n-1}$ 与 $N_{n-1}$ 独立.

(5) $P\left\{N_n - N_{n-1} = \begin{cases}1,\\0\end{cases}\right\} = \begin{cases}p,\\q.\end{cases}$

(6) $N_n$ 为简单随机徘徊，即若 $X_n \triangleq N_n - N_{n-1}$，则 $\{X_k\}$ 为 i.i.d. 随机列.

11. 证明与习题 10 反向的结论：若 $N_n$ 为简单随机徘徊，那么对于

$$S_k \triangleq \inf\{n; N_n = k\}, T_k \triangleq S_k - S_{k-1} \quad (S_0 \triangleq 0),$$

有

(1) $P(S_1 = l_1, \cdots, S_n = l_n) = p^n q^{l_n - n} \mathbf{1}_{0 < l_1 < \cdots < l_n}$；

(2) $\{T_k\}$ 为 i.i.d. 几何分布.

注：由习题 10 与习题 11 可知应有如下的对应：

| 连续时间 | 离散时间 |
| --- | --- |
| Poisson 过程 $N_t$ | 简单随机徘徊 $N_n$ |
| 1) 更新间隔：指数分布 | 几何分布 |

(2) 更新时间：$\Gamma$ 分布　　　　负二项分布

(3) "更新强度"：$\lim\limits_{t\to 0}\dfrac{P(N_t>0)}{t}=\lambda$　　　　$P(N_1>0)=p$

(4) 联合更新密度：$\lambda^n e^{-\lambda t_n} 1_{0<t_1<\cdots<t_n}$　　　　$p^n q^{t_n-n} 1_{0<t_1<\cdots<t_n}$

(5) 条件分布：

$P(N_s=k|N_t=l)$　$(s<t)$ 为二项分布

$P(N_m=k|N_n=l)$　$(m<n)$ 为超几何分布

12. 设 $(0,t]$ 内进入计数器的粒子数 $N_t$ 是参数为 $\lambda$ 的 Poisson 过程. 又若每个粒子在进入计数器后能被记录下来的概率为 $p$. 记 $X_t$ 为 $[0,t]$ 中被记录到的粒子数，求 $X_t$ 的母函数及 $EX_t$，$\mathrm{Var}X_t$.

13. 设 $N_t$ 为 Poisson 过程，求

$$P(N_s=k|N_t=n)\quad(k\leqslant n,s<t).$$

14. 设 $\{N_t;t\geqslant 0\}$，$\{M_t;t\geqslant 0\}$ 分别为相互独立的参数为 $\lambda$ 与 $\mu$ 的 Poisson 过程，求 $P(N_t=k|M_t+N_t=n)(k\leqslant n)$.

15. 把习题 13 与习题 14 移植到离散时间参数的情形.

16.（广义更新流与广义更新过程） 设 $\{X_n\}$ 为独立的正随机变量列，$S_n=X_1+\cdots+X_n$，$N_t$ 为 $(0,t]$ 出现的 $S_k$ 个数，即

$$N_t=\max\{k:S_k\leqslant t\}.$$

$N_t$ 称为由广义更新流 $\{X_n\}$ 决定的广义更新过程，求证 $N_t$ 是马氏链.

17.（具线性增长率的纯生过程） 设 $X_t$ 取非负整数值，满足

$$P(X_{t+h}-X_t=j|X_t=i)=\begin{cases}i\lambda h+o(h), & j=i+1;\\ 1-i\lambda h+o(h), & j=i;\\ o(h) & \text{其他}.\end{cases}$$

求证 $X_t$ 是时齐马氏链，而且

(1) $p_{i,i+k}(t)\triangleq P(X_t=i+k|X_0=i)$

$$=\begin{cases}\delta_{0k}, & i=0;\\ 0, & k<0,i>0\\ C_{i+k-1}^{k}(e^{-\lambda t})^i(1-e^{-\lambda t})^k, & k\geqslant 0,i>0.\end{cases}$$

(2) 若 $X_0 = m$，则 $EX_t = me^{-\mu t}$，$\mathrm{Var}X_t = me^{-2\mu t}(1 - e^{-\mu t})$.

提示：仿 Poisson 过程平稳情形，建立 $p_{m,j}(t)$ 满足的方程，归纳的逐个求解，再求条件特征函数从而得 $EX_t$，$\mathrm{Var}X_t$.

18. 若整值过程 $X_t$ 满足

$$
P(X_{t+h} = j \mid X_t = i) = \begin{cases} \mu_i h + o(h), & j = i - 1; \\ 1 - (\lambda_i + \mu_i)h + o(h), & j = i; \\ \lambda_i h + o(h), & j = i + 1; \\ o(h), & \text{其他} \end{cases}
$$

(这种过程即生灭过程)，而且存在 $C$ 使 $\forall i$ 有

$$
\lambda_i + \mu_i \leqslant C.
$$

求

$$
\lim_{T\to\infty} \frac{1}{T}\int_0^T p_{ij}(t)\mathrm{d}t.
$$

其中

$$
p_{ij}(t) \triangleq P(X_t = j \mid X_0 = i).
$$

19. 设非时齐的离散更新流 $\{X_n\}$ (即 $\{X_n\}$ 为独立列)满足

$$
X_k \sim \begin{pmatrix} 0 & 1 \\ q_k & p_k \end{pmatrix} \quad (q_k = 1 - p_k).
$$

$$
S_n = X_1 + X_2 + \cdots + X_n, \quad N_n = \max\{k; S_k \leqslant n\}.
$$

求证年龄

$$
A_n \triangleq n - S_{N_n}
$$

是马氏链，其转移阵 $(p_{ij})$ 为

$$
p_{ij} = \begin{cases} q_i, & j = 0; \\ p_i, & j = i + 1; \\ 0, & \text{其他}. \end{cases}
$$

再问什么时候 $\{A_n; n \geqslant 0\}$ 正常返(此时 $\pi = ?$)、什么时候暂态，什么时候零常返？

20. 若 $\xi$ 与 $\eta$ 是两个相互独立的 Poisson 过程，则 $\zeta = \{\zeta_t : t \geqslant 0\}$ 仍是 Poisson 过程，其中

$$\zeta_t = \xi_t + \eta_t \quad (t \geqslant 0).$$

21. 若 $\alpha_t^n \triangleq E\xi_t^n$，其中 $\xi_t$ 是 Poisson 过程，试求出 $\alpha_t^n$.

22. 试证明 Poisson 过程是非常返的.

23. 设 $\xi$ 是连续时间参数的分支过程，而且其转移概率 $(p_{ij}(t))$ 满足前进方程，试证明对 $i \geqslant 1$ 有

$$\frac{\partial g_t^{(i)}(z)}{\partial t} = \sum_{k=0}^{\infty} q_{ik}(g_t^{(i)}(z))^k,$$

其中

$$g_t^{(i)}(z) = \sum_j p_{ij}(t) z^j \quad (i \geqslant 1).$$

24.（Yule 线性纯生过程） 考虑连续参数分支过程的特例，即(5.33)中

$$f_j = 0, \quad f_2 = 1 \quad (j \neq 2)$$

的情形. 试证明这时存在唯一的转移阵 $(p_{ij}(t))$ 满足后退方程：

$$p'_{ik}(t) = \sum_j q_{ik} p_{kj}(t), \quad p_{ij}(0) = \delta_{ij}.$$

25. 设 $\xi$ 是取 $+1$ 或 $-1$ 的二值随机变量：

$$P(\xi = +1) = P(\xi = -1) = \frac{1}{2}.$$

又设 $X_t = (-1)^{N_t}$，其中 $\{N_t : t \geqslant 0\}$ 是与 $\xi$ 独立的 Poisson 过程，求证是 $X_t$ 马氏链. 求它的 Q-矩阵及平稳不变分布，并证明它是可逆的.

26. 若 $N_t$ 为 Poisson 过程，$C$ 为常数，求证 $X_t \triangleq N_t \wedge C$ 为时齐马氏链.

27. 求证 Poisson 流的年龄与余寿相互独立.

28. $\{S_n^{(i)}\}$ 为以 $\lambda_i$ 参数的相互独立的两个（即 $i = 1,2$）指数流，其对应的计数过程为 Poisson 过程 $N_t^{(i)}$，求证 $N_{S_{n+1}^{(1)}}^{(2)} - N_{S_n^{(1)}}^{(2)}$ 服

从参数为 $\dfrac{\lambda_2}{\lambda_1+\lambda_2}$ 的几何分布.

29. 美容院有 $N$ 个美发师且只能容纳 $M$ 个顾客($M \geqslant N$),第 $M+1$ 个顾客就自动离去. 设顾客以参数为 $\lambda$ 的指数到达流,而美容所需时间服从参数为 $\mu$ 的指数分布而且与到达情形独立. 求时刻 $t$ 在美容院中顾客数 $N_t$ 的平稳分布.

30. 有 $m$ 台设备,每台的正常运转时间都服从参数为 $\lambda$ 的指数分布,检修所需时间均服从参数为 $\mu$ 的指数分布且与运转情形独立,各台设备也独立. 假定采用如下的检修方案:有 $k$ 台不能正常运转就全部停工检修. 记时刻 $t$ 不能正常运转的设备数为 $N_t$:

(i) 求证 $N_t$ 为具有下述转移速率阵的 Q-过程:

$$Q=(q_{ij}),\quad 0\leqslant i,j\leqslant k,$$

$$q_{ij}=\begin{cases}-(m-i)\lambda, & 0\leqslant j=i<k;\\ (m-i)\lambda, & 0\leqslant i<k,\ j=i+1;\\ \mu, & i=k,\ j=0;\\ -\mu, & i=j=k;\\ 0, & \text{其他情形}.\end{cases}$$

(ii) 求 $N_t$ 的平稳分布.

(iii) 令 $m_k=m-\lim\limits_{t\to\infty}EN_t$,证明:

$$m_k\geqslant m_{k+1}\Rightarrow m_k\geqslant m_{k+i}\quad (i\geqslant 1),$$

并讨论什么样的 $k$ 使 $m_k$ 最大.

31. 在 $M/M/2$ 中设顾客流的指数参数为 $\lambda$,服务时间分布的指数参数为 $\mu$,且 $\dfrac{\lambda}{2\mu}<1$. 如果改用一个比原来的两条服务线快一倍的一条服务线(即改用 $M/M/1$,但是服务时间的指数分布为 $2\mu$). 试以比较这两种服务系统在稳定情形下的平均排队的队伍长度判断其优劣.

32. 在 $M/M/1$ 中若每当队伍长度 $> N$ 时就增加一条服务线，而当队伍长度 $\leqslant N$ 时就撤掉新增加的那条服务线. 求排队队伍长度 $N_t$ 的平稳分布(队伍长度是指在系统中顾客的数目).

33.(带随机消失的 $M/M/1$ ) 在下列情形下分别求在时刻 $t$ 系统中顾客数 $N_t$ 的平稳分布：

(i) 新来的顾客看到原来的 $N_t > l(l > 1)$ 则以概率 $p$ 留下，以概率 $1-p$ 离去.

(ii) 新来的顾客看到原系统中 $N_t > l$ 则离去，而在 $N_t \leqslant l$ 时则以概率 $p = \frac{l - N_t}{l(N_t + 1)}$ 留下(这时 $N_t$ 是一个具体的数)，以概率 $1-p$ 离去.

(iii) 在(ii)中改为 $p = \frac{N_t}{1 + N_t}$.

34. 出租车站以指数流 $\lambda$ 来车，顾客则以指数流 $\mu$ 独立的到站，假定一辆车只载一个客人，且顾客到站时如看到已有两人在等候时就自动离开. 设 $\frac{\lambda}{\mu} < 1$，求等候客人的出租车在稳定情形下的平均数.

35. 分别在 $GI/M/N$,$GI/M/\infty$ 中把输入流的计数过程由 Poisson 过程改为复合 Poisson 过程，说明此时在时刻 $t$ 在系统中的顾客数 $N_t$ 仍为马氏链，写出它的转移速率阵 $Q$，并讨论什么时候有平稳分布，它是什么.

# 第六章　随机迭代映射与离散时间连续状态的马氏链

在许多实际问题中，有时候时间是离散的，例如在计算机科学中，由于计算总是一步步地进行的，时间指标自然应是离散的. 此外，很多情况下，观察是定时的，例如每秒一次，每分一次，每小时一次，等等，于是采样时间自然只能也是离散的了. 另一方面，我们又不宜将状态作为离散的来处理. 例如求解非线性方程

$$x^4 + \operatorname{tg} x = 0 \tag{6.0}$$

时，如果我们把方程(6.0)的解看成是达到函数

$$f(x) = (x^4 + \operatorname{tg} x)^2$$

的最小值(零)的点，而仿照前面提到的模拟退火的思想去求解，那么 $X$ 就不宜看成是离散的，而求解的过程遇到一列具有类似于第二章中讲过的"马氏性"而取值连续的随机变量序列，或者说，状态连续，时间离散的马氏链. 本章对连续状态马氏链作一个初步的研究.

## §6.1　随机迭代映射与连续状态马氏链

首先让我们来看一些例子.

**例 6.1**　设 $W_n$ 是 i.i.d. 随机变量序列，而

$$\xi_{n+1} = f(\xi_n) + W_n,$$

并且 $\xi_n$ 与 $W_n$ 独立. 那么$\{\xi_n: n\geqslant 0\}$是随机迭代映射所决定的随机序列. 这样的随机序列十分常见，例如在一个迭代记录中，每次会有误差$\{W_n\}$，就会有上述结果. 在 $\xi_n = x$ 为已知的条件下，我们有 $\xi_{n+1} = f(x) + W_n$，因而它与$(\xi_0, W_1, \cdots, W_{n-1})$独立，从而也与$(\xi_n,$

$\zeta_{n-1},\cdots,\zeta_0$)独立. 设 $W_n$ 具有分布密度 $\rho(y)$，于是

$$
\begin{aligned}
P(\zeta_{n+1} &\leqslant y \mid \zeta_n = x, \zeta_{n-1} = x_{n-1}, \cdots, \zeta_0 = x_0) \\
&= P(f(x) + W_n \leqslant y \mid \zeta_n = x, \zeta_{n-1} = x_{n-1}, \cdots, \zeta_0 = x_0) \\
&= P(f(x) + W_n \leqslant y) = P(W_n \leqslant y - f(x)) \\
&= \int_{-\infty}^{y-f(x)} \rho(z)\mathrm{d}z.
\end{aligned}
$$

同样也有

$$
P(\zeta_{n+1} \leqslant y \mid \zeta_n = x) = \int_{-\infty}^{y-f(x)} \rho(z)\mathrm{d}z = \int_{-\infty}^{y} \rho(z - f(x))\mathrm{d}z.
$$

可见我们有

$$
\begin{aligned}
P(\zeta_{n+1} \leqslant y \mid \zeta_n = x, \zeta_{n-1} = x, \cdots) &= P(\zeta_{n+1} \leqslant y \mid \zeta_n = x) \\
&= \int_{-\infty}^{y-f(x)} \rho(z)\mathrm{d}z = \int_{-\infty}^{y} \rho(z - f(x))\mathrm{d}z.
\end{aligned}
$$

第一个等式与第三章中的马氏性质非常相似，不同处在于第三章中的 $\zeta_n$ 是取可数值的，而现在的 $\zeta_n$ 是取连续值的. 这种随机序列称为连续状态的马氏链，而

$$
F(x,y) \triangleq \int_{-\infty}^{y} \rho(z - f(x))\mathrm{d}z (= P(\zeta_{n+1} \leqslant y \mid \zeta_n = x)).
$$

在 $x$ 固定的条件下是 $y$ 为变量的一个分布函数，称为从 $\zeta_n$ 转移到 $\zeta_{n+1}$ 的转移分布函数，它具有密度 $\rho(y-f(x))$，称为从 $\zeta_n$ 到 $\zeta_{n+1}$ 的转移概率密度，这是一个条件概率密度.

下面我们来看一个稍复杂一些的例子：

**例 6.2** 设 $\{W_n\}$ 为独立序列，且 $W_n$ 有分布密度 $\rho_n(y)$，

$$
\zeta_{n+1} = f(\zeta_n, W_n),
$$

$\zeta_0$ 与 $\{W_n\}$ 独立. 此时 $W_n$，$f(x,W_n)$ 应与 $\{\zeta_0,\zeta_1,\cdots,\zeta_n\}$ 独立，于是

$$
\begin{aligned}
P(\zeta_{n+1} &\leqslant y \mid \zeta_n = x, \zeta_{n-1} = x_{n-1}, \cdots, \zeta_0 = x_0) \\
&= P(f(x,W_n) \leqslant y \mid \zeta_n = x, \zeta_{n-1} = x_{n-1}, \cdots, \zeta_0 = x_0) \\
&= P(f(x,W_n) \leqslant y) = \int_{f(x,z) \leqslant y} \rho_n(z)\mathrm{d}z,
\end{aligned}
$$

同样也有

$$P(\zeta_{n+1}\leqslant y|\zeta_n=x)=\int_{f(x,z)\leqslant y}\rho_n(z)\mathrm{d}z.$$

记

$$F_n(x,y)\triangleq\int_{f(x,z)\leqslant y}\rho_n(z)\mathrm{d}z,$$

那么在 $\zeta_n=x$ 已知的条件下，$\zeta_{n+1}$的条件分布函数为 $F_n(x,y)$.

在这两个例子中，我们都有

$P(\zeta_{n+1}\leqslant y|\zeta_n=x,\zeta_{n-1}=x_{n-1},\cdots)=P(\zeta_{n+1}\leqslant y|\zeta_n=x)$.

所以$\{\zeta_n;n\geqslant 0\}$都是连续状态的马氏链．但是在例 6.1 中上式右方 $=\int_{-\infty}^{y-f(x)}\rho(z)\mathrm{d}z$，与 $n$ 无关，所以称为时齐的马氏链．而在例 6.2 中上式右方$\int_{f(x,z)\leqslant y}\rho_n(z)\mathrm{d}z$，它依赖于 $n$，所以$\{\zeta_n\}$是非时齐的马氏链.

从这两个例子我们可以抽象成下述定义：

**定义 6.1（连续状态马氏链）** 随机过程$\zeta\triangleq\{\zeta_n,n\geqslant 0\}$称为连续状态离散（时间）参数的马氏链，如果对$\forall\ m\geqslant 1,n\geqslant 0$ 及 $x,y$，$x_0,\cdots x_{n-1}\in R$，都有

$$\begin{aligned}&P(\zeta_{n+m}\leqslant y|\zeta_n=x,\zeta_{n-1}=x_{n-1},\cdots,\zeta_0=x_0)\\&\quad=P(\zeta_{n+m}\leqslant y|\zeta_n=x).\end{aligned}\tag{6.1}$$

为了回避测度论，我们假定$(\zeta_0,\cdots,\zeta_n,\zeta_{n+m})$有联合密度 $f$，这时(6.1)式理解为

$$\begin{aligned}&\int_{-\infty}^{y}f(x_0,\cdots,x_{n-1},x,z)\mathrm{d}z\left[\int f(x_0,\cdots,x_{n-1},x,z)\mathrm{d}z\right]^{-1}\\&\quad=\frac{\int_{-\infty}^{y}\left[\int\cdots\int f(x_0,\cdots,x_{n-1},x,z)\mathrm{d}x_0\cdots\mathrm{d}x_{n-1}\right]\mathrm{d}z}{\iint\cdots\int f(x_0,\cdots x_{n-1},x,z)\mathrm{d}x_0\cdots\mathrm{d}x_{n-1}\mathrm{d}z}.\end{aligned}\tag{6.1$'$}$$

记$(\zeta_n,\zeta_{n+m})$的联合密度为 $f_{n,m}(x,y)$，于是(6.1) 的右方有“分布密度”$f_{n,m}(x,y)\left[\int f_{n,m}(x,y)\mathrm{d}y\right]^{-1}$，称为从 $\zeta_n$ 到 $\zeta_{n+m}$的**转移概率**

**密度**（条件概率密度），记为 $p(n,x;m,y)$.

连续状态马氏链有与 § 3.1 类似的许多性质：

$$\int p(n,x;m,y)\mathrm{d}y = 1, \tag{6.3}$$

$$\int p(n,x;m,z)p(m,z;l,y)\mathrm{d}z = p(n,x;l,y) \quad (n < m < l) \tag{6.4}$$

（形式上，上式就是在命题 3.1 中将 $i,j,k$ 换成 $x,y,z$，对 $j$ 求和换成对 $y$ 积分.（6.4）称为 Kolmogorov-Chapman 方程）.

**定义 6.2** 连续状态马氏链 $\zeta=\{\zeta_n:n\geqslant 0\}$ 称为时齐的，如果（6.1）右方不依赖 $n$.

我们仍考虑有密度情形，这样对时齐情形我们简记

$$p^{(m)}(x,y) \triangleq p(0,x;m,y),$$
$$p(x,y) \triangleq p(0,x;1,y) = p(n,x;n+1,y). \tag{6.5}$$

于是（6.4）化为

$$p^{(m+1)}(x,y) = \int p(x,z)p^{(m)}(z,y)\mathrm{d}z\ . \tag{6.6}$$

此外，与离散状态马氏链类似，我们在本章以下部分永远假定有转移密度 $\{p(n,x;m,y)\}$，而且满足（6.3），（6.4），但是让初值 $\zeta_0$ 可以取任一值或任一分布.

利用（6.3），（6.4），还可与离散状态马氏链类似地有：如果 $\zeta_0$ 有分布密度 $\rho_0(y)$，那么

$$\begin{aligned}
&P(\zeta_0\in\Gamma_0,\zeta_1\in\Gamma_1,\cdots,\zeta_n\in\Gamma_n)\\
&= \int\cdots\int_{x_0\in\Gamma_0,\cdots,x_n\in\Gamma_n} \rho_0(x_0)p(0,x_0;1,x_1)\cdots\\
&\quad\times p(n-1,x_{n-1};n,x_n)\mathrm{d}x_0\cdots\mathrm{d}x_n.
\end{aligned}$$

但是对这个公式作严格的证明需要用测度论的工具.

**例 6.3** 设 $W_n$ 为 i.i.d. 的 $N(0,1)$（标准正态）随机序列，又

$f(x)\geqslant 0$ 且二阶光滑，再设 $f$ 的全部最小值点组成集合 $S=\{x_1,\cdots,x_l\}$（从而 $\nabla f(x_k)=0, k\leqslant l$），考虑

$$\zeta_{n+1}=\zeta_n+h\nabla f(\zeta_n)+\epsilon_n W_n f(\zeta_n)\quad (h>0),\quad \zeta_0=x_0.$$

这是例 6.2 的特例，而且在 $\zeta_n=x$ 已知这一条件下到 $\zeta_{n+1}$ 的转移函数

$$F(x;y)=\int_{x+h\nabla f(x)+\epsilon_n f(x)z\leqslant y}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}\mathrm{d}z$$

$$=\int_{-\infty}^{\frac{y-x-h\nabla f(x)}{\epsilon_n f(x)}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}\mathrm{d}z,$$

它具有转移密度

$$p(n,x;n+1,y)=\frac{1}{\epsilon_n f(x)}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}\left(\frac{y-x-h\nabla f(x)}{\epsilon_n f(x)}\right)^2}$$

$$=N(x+h\nabla f(x),\epsilon_n^2 f(x)^2)(y).$$

所以 $\zeta_n$ 是非时齐的连续状态马氏链，非时齐是由“小参数”$\epsilon_n$ 依赖于 $n$ 引起的.

我们注意如果 $\zeta_n(w)\in S$，就只有小扰动项 $\epsilon_n W_n f(\zeta_n)$. 也就是这个链的轨道一旦到达了 $f$ 的最小值再离开它的小邻域的可能性就较小了. 这个思想启示我们思考是否能用这个链来“近似”地找到 $f$ 的最小值的位置，以后我们会介绍，如果适当地选取 $\epsilon_n\to 0$，（例如 $\epsilon_n=c/\log(n+1)$，$c$ 足够大），那么有 $P(\zeta_n\in B_\delta(S))\xrightarrow{n\to\infty}1$（对任意固定 $\delta>0$，其中 $B_\delta(S)$ 是指 $S$ 中所有点的 $\delta$ 邻域的并）. 也就是迭代足够多次后以近似于 1 的概率 $\zeta_n$ 会到达 $f$ 的取最小值的集合任意近处. 这就是用随机迭代模型来作模拟退火的基本思想，这里 $\epsilon_n$ 依赖 $n$ 是至关重要的. 这一数学模型对于在计算机上作涉及随机因素或噪声的大部分运算是一个很好的表达公式. 因而它具有广泛的应用领域.

求函数 $g(x)$ 零点的问题可以化为求 $g(x)^2$ 的最小值问题，仍可用模拟退火去作.

## §6.2 Dobrushin 不等式、指数遍历性与收敛性

在第三章中，我们曾对离散马氏链的在 $t\to\infty$ 时的极限行为进行了讨论. 在本节中，我们将利用 Dobrushin 不等式，对于连续状态的马氏链得到与定理 3.8 相应的结果. 这类结果要求转移概率密度存在，并满足较强的正则条件，但是其结论不仅可以得到 $t$ 步转移概率密度当 $t\to\infty$ 时的极限趋于唯一的不变概率分布密度，而且保证误差具有指数型的下降速度(称为**指数遍历性**). 这一结论比之于一般的马氏链的遍历定理强得多.

### 1. Dobrushin 不等式

设 $\rho_i(x)(i=1,2)$ 为 $R$(也可以是 $R^d$)上分布密度，$p(x,y)$ 是 $R$ 上的转移密度，即

$$p(x,y)\geqslant 0,\quad \int p(x,y)\mathrm{d}y=1, \tag{6.7}$$

定义函数 $v$ 的模为

$$\|v\|=\int|v(x)|\mathrm{d}x.$$

即 $v$ 的 $L_1$ 模. 当 $v$ 为两个分布密度差时，$\|v\|$ 即为相应的两个分布函数的差的"全变差".

**命题 6.1 (Dobrushin 不等式)** 在前面的假定下，我们有

$$\begin{aligned}&\left\|\int p(x,\cdot)\rho_1(x)\mathrm{d}x-\int p(x,\cdot)\rho_2(x)\mathrm{d}x\right\|\\&\leqslant C(p)\|\rho_1-\rho_2\|,\end{aligned} \tag{6.8}$$

其中 $C(p)$(称为转移密度 $p(x,y)$ 的 Dobrushin 数)由下式所定义

$$C(p)\triangleq\frac{1}{2}\sup_{x,y}\int|p(x,z)-p(y,z)|\mathrm{d}z. \tag{6.9}$$

**证明** 记

$$\rho_+(x)=(\rho_1(x)-\rho_2(x))^+\Big/\frac{1}{2}\|\rho_1-\rho_2\|,$$

那么由于

$$\int[(\rho_1-\rho_2)^+-(\rho_1-\rho_2)^-]=\int(\rho_1-\rho_2)=\int\rho_1-\int\rho_2=0,$$

$$\int[(\rho_1-\rho_2)^++(\rho_1-\rho_2)^-]=\int|\rho_1-\rho_2|=\|\rho_1-\rho_2\|.$$

我们得到

$$\int\rho_+=\int\rho_-=\frac{2}{\|\rho_1-\rho_2\|}\int(\rho_1-\rho_2)^+=\frac{\frac{1}{2}\|\rho_1-\rho_2\|}{\frac{1}{2}\|\rho_1-\rho_2\|}=1,$$

同时我们有

$$\rho_1-\rho_2=(\rho_1-\rho_2)^+-(\rho_1-\rho_2)^-=\frac{1}{2}(\rho_+-\rho_-)\|\rho_1-\rho_2\|.$$

于是

$$\begin{aligned}
&\left\|\int\rho_1(x)p(x,\cdot)\mathrm{d}x-\int\rho_2(x)p(x,\cdot)\mathrm{d}x\right\|\\
&=\int\left|\int(\rho_1(x)-\rho_2(x))p(x,z)\mathrm{d}x\right|\mathrm{d}z\\
&=\frac{1}{2}\|\rho_1-\rho_2\|\int\left|\int(\rho_+(x)-\rho_-(x))p(x,z)\mathrm{d}x\right|\mathrm{d}z\\
&=\frac{1}{2}\|\rho_1-\rho_2\|\int\left|\iint\rho_+(x)\rho_-(y)(p(x,z)-p(y,z))\mathrm{d}x\mathrm{d}y\right|\mathrm{d}z\\
&\leqslant\|\rho_1-\rho_2\|\iint\rho_+(x)\rho_-(y)\left(\int\frac{1}{2}|p(x,z)-p(y,z)|\mathrm{d}z\right)\mathrm{d}x\mathrm{d}y\\
&\leqslant C(p)\|\rho_1-\rho_2\|\iint\rho_+(x)\rho_-(y)\mathrm{d}x\mathrm{d}y=C(p)\|\rho_1-\rho_2\|.
\end{aligned}$$

**例 6.4** 若 $\{W_n\}$ 独立都服从 $N(0,1)$，$f(x)$ 为非常数的有界连续函数，考虑连续状态时齐马氏链 $\{\zeta_n,n\geqslant0\}$：

$$\zeta_{n+1}=f(\zeta_n)+W_n. \tag{6.10}$$

于是其转移密度为 $p(x,z)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(z-f(x))^2}{2}}$. 对于使 $f(x)\neq f(y)$ 的点 $(x,y)$，记方程 $\mathrm{e}^{-\frac{(z-f(x))^2}{2}}=\mathrm{e}^{-\frac{(z-f(y))^2}{2}}$ 的唯一解为

$$z=h(x,y)\triangleq\frac{f(x)+f(y)}{2}.$$

由于 $f(x)$ 是有界连续函数，可见 $h(x,y)$ 为有界函数. 我们有：当 $f(x)<f(y)$ 时，

$$
\begin{aligned}
&\int |p(x,z)-p(y,z)|\mathrm{d}z \\
&= \int_{-\infty}^{h(x,y)} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-f(x))^2}{2}}\mathrm{d}z + \int_{h(x,y)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-f(y))^2}{2}}\mathrm{d}z \\
&= \int_{-\infty}^{h(x,y)-f(x)} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\mathrm{d}u + \int_{h(x,y)-f(y)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\mathrm{d}u \\
&\leqslant \int_{-\infty}^{m_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\mathrm{d}u + \int_{m_1}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\mathrm{d}u \\
&= \Phi(m_2) + (1-\Phi(m_1)),
\end{aligned}
$$

其中

$$
m_1 \triangleq \inf_{x\neq y}(h(x,y)-f(x)),
$$

$$
m_2 \triangleq \sup_{x\neq y}(h(x,y)-f(y)),
$$

$$
\Phi(\alpha) \triangleq \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\alpha} e^{\frac{-u^2}{2}}\mathrm{d}u.
$$

由 $f$ 的有界性，可知 $-\infty<m_1<m_2<\infty$，故而可取

$$
\begin{aligned}
C(p) &= \frac{1}{2}\sup_{x,y}\frac{1}{\sqrt{2\pi}}\int\left|e^{\frac{-(z-f(x))^2}{2}} - e^{\frac{-(z-f(y))^2}{2}}\right|\mathrm{d}z \\
&\leqslant \frac{1}{2}(1-\Phi(m_1)+\Phi(m_2)) < 1.
\end{aligned}
$$

**例 6.5** 若连续状态马氏链 $\{\zeta_n, n\geqslant 0\}$ 具有满足以下条件的转移密度 $p(x,y)$：

$$
p(x,y) \geqslant g(y) \geqslant 0 \quad (\forall\ x), \tag{6.11}
$$

那么此链的 Dobrushin 常数

$$
C(p) \leqslant 1 - \int g(y)\mathrm{d}y.
$$

注意 $0\leqslant \int g(y)\mathrm{d}y \leqslant \int p(x,y)\mathrm{d}y = 1$.

**证明** 由 Dobrushin 不等式的证明可知：

$$
\begin{aligned}
C(p) &= \frac{1}{2}\sup_{x,y}\int |p(x,z) - p(y,z)|\mathrm{d}z \\
&= \sup_{x,y}\int_{p(x,z)\geqslant p(y,z)} (p(x,z) - p(y,z))\mathrm{d}z \\
&\leqslant \sup_{x,y}\int_{p(x,z)\geqslant p(y,z)} (p(x,z) - g(z))\mathrm{d}z \\
&\leqslant \sup_{x,y}\int (p(x,z) - g(z))\mathrm{d}z = 1 - \int g(z)\mathrm{d}z .
\end{aligned}
$$

**定义 6.3** 具有转移密度 $p(x,y)$ 的连续状态时齐马氏链 $\{\zeta_n, n\geqslant 0\}$ 称为**指数 $L_1$ 遍历(或强指数遍历的)**,如果∃ $\rho^{(\infty)}(y)$, $n_0$ 及 $c>0$, $0<r<1$,使得∀ $x$ 有

$$
\int |p^{(n)}(x,y) - \rho^{(\infty)}(y)|\mathrm{d}y < cr^n \quad (n > n_0), \qquad (6.12)
$$

其中 $p^{(n)}(x,y)$ 是 $n$ 步转移密度:

$$
p^{(n)}(x,y) \triangleq \int p(x,z)p^{(n-1)}(z,y)\mathrm{d}z,
$$

$$
p^{(1)}(x,y) \triangleq p(x,y). \qquad (6.13)
$$

**定义 6.4** 连续状态的时齐马氏链 $\{\zeta_n, n\geqslant 0\}$ 称为**指数遍历的**,如果∃ 分布函数 $F^{(\infty)}$, $n_0$, $c>0$, $0<r<1$,使对∀ $x, y$ 有

$$
|P(\zeta_n(x) \leqslant y) - F^{(\infty)}(y)| \leqslant cr^n \quad (n \geqslant n_0), \qquad (6.14)
$$

其中 $\zeta_n(x)$ 是初值 $\zeta_0 = x$ 的马氏链 $\zeta_n$.

显见强指数遍历一定指数遍历. $F^{(\infty)}$ 称为 $\zeta_n$ 的极限分布,也称为 $\zeta_n$ 的平稳分布,以此为初分布,就得到一个连续状态的平稳马氏链.

**定理 6.2** 若具有转移密度的连续状态时齐马氏链 $\zeta_n$ 的 Dobrushin 常数 $C(p)<1$,则马氏链指数 $L_1$ 遍历.

**证明梗概**(初读时可略去)

1° 对任意分布密度 $\rho_1, \rho_2$,利用 Dobrushin 不等式有

$$
\left\| \int \rho_1(x)p^{(n)}(x,\cdot)\mathrm{d}x - \int \rho_2(x)p^{(n)}(x,\cdot)\mathrm{d}x \right\|
$$

$$
= \left\| \iint p^{(n-1)}(x,y)p(y,\cdot)\mathrm{d}y(\rho_1(x) - \rho_2(x))\mathrm{d}x \right\|
$$

$$= \left\| \int p(y,\cdot) \left[ \int p^{(n-1)}(x,y)(\rho_1(x) - \rho_2(x))\mathrm{d}x \right] \mathrm{d}y \right\|$$

$$\leqslant C(p) \left\| \int \rho_1(x) p^{(n-1)}(x,\cdot)\mathrm{d}x - \int \rho_2(x) p^{(n-1)}(x,\cdot)\mathrm{d}x \right\|$$

$$\leqslant \cdots \leqslant (C(p))^n \| \rho_1 - \rho_2 \|.$$

2° 可以证明(例如用压缩映象原理)存在 $\rho^{(\infty)}$,使对任意分布密度 $\rho$ 有

$$\left\| \int \rho(x) p^{(n)}(x,\cdot)\mathrm{d}x - \rho^{(\infty)} \right\| \leqslant \| \rho - \rho^{(\infty)} \| (C(p))^n.$$

3° 所以 $\rho^{(\infty)}$ 是分布密度,而且具有如下的"不变性":

$$\rho^{(\infty)}(y) = \int p(x,y)\rho^{(\infty)}(x)\mathrm{d}x.$$

4° $\int |p^{(n)}(x,y) - \rho^{(\infty)}(y)|\mathrm{d}y = \| p^{(n)}(x,\cdot) - \rho^{(\infty)} \|$

$$= \left\| \int p(x,z) p^{(n-1)}(z,y)\mathrm{d}z - \rho^{(\infty)}(y) \right\|,$$

在 2°中把 $p(\cdot, x)$ 看成 $\rho(x)$,并把 $x$ 换成 $z$,可见

$$\| p(x,\cdot) - \rho^{(\infty)} \| (C(p))^{n-1} \leqslant 2(C(p))^{n-1}.$$

这是因为有

$$\| p(x,\cdot) - \rho^{(\infty)} \| = \int |p(x,y) - \rho^{(\infty)}(y)|\mathrm{d}y$$

$$\leqslant \int p(x,y)\mathrm{d}y + \int \rho^{(\infty)}(y)\mathrm{d}y = 2.$$

**推论** 例 6.5 中连续状态马氏链当 $\int g(y)\mathrm{d}y > 0$ 时是指数 $L^1$ 遍历的.

**定理 6.3** 设 $\{W_n\}$ 为 i. i. d. 随机列,$f(u,v)$ 连续,且关于 $u$ 单调不减,再设

$$\begin{cases} \zeta_0 = x, \\ \zeta_{n+1}(x) = f(\zeta_n(x), W_n). \end{cases} \tag{6.15}$$

又如果存在 $z_0$ 使

$$P(\sup_x \zeta_1(x) \leqslant z_0), P(\inf_x \zeta_1(x) \geqslant z_0) \geqslant \delta > 0, \tag{6.16}$$

则连续状态马氏链$\{\zeta_n:\forall x, n\geqslant 0\}$指数遍历(我们在$W_n$只取有限个值的情形给出证明梗概，对一般情形若掌握了测度论的知识，就不难推广此证明).

**证明梗概**

1° (1) 首先注意：对$\forall x, y\in R$，当$z\geqslant z_0$时有

$$
\begin{aligned}
\{\zeta_1(x)\leqslant z\} &\subset \{\zeta_1(y)\leqslant z\}\cup\{\zeta_1(x)\leqslant z, \zeta_1(y)>z\} \\
&\subset \{\zeta_1(y)\leqslant z\}\cup\{\sup_u \zeta_1(u)>z_0\};
\end{aligned}
$$

类似地，当$z<z_0$时有

$$
\{\zeta_1(x)\leqslant z\}\subset\{\zeta_1(y)\leqslant z\}\cup\{\inf_u \zeta_1(u)<z_0\}.
$$

于是

$$
\{\zeta_1(x)\leqslant z\}\setminus\{\zeta_1(y)\leqslant z\}\subset\{\inf_u \zeta_1(u)<z_0\},
$$

$$
\begin{aligned}
&|P(\zeta_1(x)\leqslant z)-P(\zeta_1(y)\leqslant z)| \\
&\quad\leqslant P(\sup_u \zeta_1(u)>z_0)\vee P(\inf_u \zeta_1(u)<z_0) \\
&\quad\leqslant 1-\delta,
\end{aligned}
$$

记

$$
\Delta_n=\sup_{x,y,z}|P(\zeta_n(x)\leqslant z)-P(\zeta_n(y)\leqslant z)|,
$$

那么$\Delta_1\leqslant 1-\delta$. 下面我们证明$\Delta_n\leqslant(1-\delta)^n$. 为此令

$$
\tilde{\zeta}_1(x)=f(x, W_n).
$$

注意$\zeta_n(x)$只依赖$W_1,\cdots,W_{n-1}$，所以$\tilde{\zeta}_1(x)$与$\{\zeta_n(x)\}$是相互独立的，而且$\tilde{\zeta}_1(x)$与$\zeta_1(x)$同分布. 利用

$$
\zeta_{n+1}(x)=f(\zeta_n(x), W_n)=\tilde{\zeta}_1(\zeta_n(x)),
$$

我们来估计

$$
\begin{aligned}
&|P(\zeta_{n+1}(x)\leqslant z)-P(\zeta_{n+1}(y)\leqslant z)| \\
&\quad=|E(I_{\{\zeta_{n+1}(x)\leqslant z\}}-I_{\{\zeta_{n+1}(y)\leqslant z\}})| \\
&\quad=|E(I_{\{\tilde{\zeta}_1(\zeta_n(x))\leqslant z\}}-I_{\{\tilde{\zeta}_1(\zeta_n(y))\leqslant z\}})|\triangleq I_1,
\end{aligned}
$$

其中$I_A(\omega)$是集合$A$的示性函数.

$$I_A(\omega)=\begin{cases}1, & \omega\in A;\\ 0, & \omega\overline{\in} A.\end{cases}$$

我们分两种情形讨论：

（1）如果 $z\leqslant z_0$，那么

$$\{\tilde{\zeta}_1(\zeta_n(x))\leqslant z\}\setminus\{\tilde{\zeta}_1(\zeta_n(y))\leqslant z\}\subset\{\inf_u\tilde{\zeta}_1(u)\leqslant z_0\}.$$

于是

$$I_1=|E[(I_{\{\tilde{\zeta}_1(\zeta_n(x))\leqslant z\}}-I_{\{\tilde{\zeta}_1(\zeta_n(y))\leqslant z\}})I_{\{\inf\limits_u\tilde{\zeta}_1(u)\leqslant z_0\}}]|.$$

再注意到 $\tilde{\zeta}_1(\zeta_n(x))=f(\zeta_n(x),W_n)$，令在 $v$ 固定时 $f(u,v)$ 关于 $u$ 的反函数记为 $f^{-1}(\cdot,v)$. 这样我们有

$$\{\tilde{\zeta}_1(\zeta_n(x))\leqslant z\}=\{\zeta_n(x)\leqslant f^{-1}(z,W_n)\},$$

所以

$$I_1=|E[(I_{\{\zeta_n(x)\leqslant f^{-1}(z,W_n)\}}-I_{\{\zeta_n(y)\leqslant f^{-1}(z,W_n)\}})I_{\{\inf\limits_u\tilde{\zeta}_1(u)\leqslant z_0\}}]|.$$

下面我们引用两个初等概率论的事实，读者应把它们作为习题证明.

1）若 $(\zeta,\eta)$ 为只取有限个值的随机向量，则

$$E(g(\eta)h(\zeta,\eta))=E(g(\eta)E(h(\zeta,\eta)|\eta)),$$

其中 $E(h(\zeta,\eta)|\eta)$ 为随机变量，它的定义为

$$E(h(\zeta,\eta)|\eta)=E(h(\zeta,\eta)|\eta=y_j)\quad(\text{当 }\eta=y_j\text{ 时}).$$

2）若还有 $\zeta,\eta$ 相互独立，则

$$E(h(\zeta,\eta)|\eta)=[E(h(\zeta,y))]_{y=\eta}.$$

在上式中取 $\zeta=(W_1,\cdots,W_{n-1})$，$\eta=W_n$，我们得到

$$\begin{aligned}I_1&=|EI_{\{\inf\limits_u\tilde{\zeta}_1(u)\leqslant z_0\}}E[(I_{\{\zeta_n(x)\leqslant f^{-1}(z,W_n)\}}-I_{\{\zeta_n(y)\leqslant f^{-1}(z,W_n)\}})|W_n]|\\ &\leqslant|EI_{\{\inf\limits_u\tilde{\zeta}_1(u)\leqslant z_0\}}[E(I_{\{\zeta_n(x)\leqslant f^{-1}(z,v)\}}-I_{\{\zeta_n(y)\leqslant f^{-1}(z,v)\}})]_{v=W_n}|\\ &\leqslant|EI_{\{\inf\limits_u\tilde{\zeta}_1(u)\leqslant z_0\}}|\Delta_n\leqslant(1-\delta)\Delta_n,\end{aligned}$$

从而得到 $\Delta_{n+1}\leqslant(1-\delta)\Delta_n$.

（2）如果 $z\geqslant z_0$，只要用 $\{\sup\limits_u\tilde{\zeta}_1(u)\leqslant z_0\}$ 来代替（1）中的

$\{\inf\limits_{u}\zeta_1(u)\leqslant z_n\}$，就可以完全相仿地证明 $\Delta_{n+1}\leqslant(1-\delta)\Delta_n$ 仍然正确.

从 $\Delta_{n+1}\leqslant(1-\delta)\Delta_n$，我们立刻可以得到 $\Delta_n\leqslant(1-\delta)^n$.

2° 由 1°并利用连续型随机变量的全概公式，不难证明 $P(\xi_n(x)\leqslant z)$ 是 Cauchy 列，而且对 $z$ 为一致收敛，其极限仍是一个分布函数而且不依赖 $x$，记这个极限为 $F^{(\infty)}(z)$. 再用 1°便得

$$\sup_{x,z}|P(\xi_n(x)\leqslant z)-F^{(\infty)}(z)|\leqslant(1-\delta)^n.$$

**注 1**：本定理条件可改为：∃ $z_0, n_0$，使

$$P(\sup\xi_{n_0}(x)\leqslant z_0),P(\inf\xi_{n_0}(x)\geqslant z_0)\geqslant\delta>0, \tag{6.17}$$

结论仍成立. 证明只需略作修改.

**注 2**：类似地，定理 6.2 也可推广为：若连续状态马氏链 $\{\xi_n:n\geqslant 0\}$ 的转移密度 $p(x,y)$ 满足：∃ $n_0$，使 $C(p^{(n_0)})<1$，那么此链仍是指数 $L^1$ 遍历的，其中 $p^{(n)}(x,y)$ 的定义为

$$p^{(2)}(x,y)\triangleq\int p(x,z)p(z,y)\mathrm{d}z,$$

$$p^{(3)}(x,y)\triangleq\int p(x,z)p^{(2)}(z,y)\mathrm{d}z,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$p^{(n+1)}(x,y)\triangleq\int p(x,z)p^{(n)}(z,y)\mathrm{d}z.$$

于是我们还有与定理 1 的推论相类似的推广：若连续状态马氏链 $\{\xi_n:n\geqslant 0\}$ 具有转移密度且满足：∃ $n_0$，使 $p^{(n_0)}(x,y)\geqslant g(y)\geqslant 0$，且 $\int g(y)\mathrm{d}y>0$，那么此链仍是指数 $L^1$ 遍历的.

**例 6.6** 设 $f$ 为有界单调函数，$\{W_n\}$ 为 i.i.d. 列，且具有严格正的连续的分布密度 $p(x)$，那么由如下定义的迭代列 $\xi_n$ 所决定的连续状态马氏链是指数 $L^1$ 遍历的：

$$\xi_{n+1}=f(\xi_n)+W_n.$$

**证明** 不妨设 $a\leqslant f(x)\leqslant b(\forall\ x)$，我们有

$$\begin{aligned}&P(\xi_{n+1}\leqslant y|\xi_n=x,\xi_{n-1}=x_1,\cdots)\\&\quad=P(W_n+f(\xi_n)\leqslant y|\xi_n=x)=P(w_n\leqslant y-f(x))\end{aligned}$$

$$= \int_{-\infty}^{y-f(x)} p(z)\mathrm{d}z.$$

它具有密度($x$ 固定,对 $y$ 求导数)$p(y-f(x))$. 于是

$$p(x,y) = p(y-f(x)) \geqslant \min_{a \leqslant z \leqslant b} p(y-z) > 0.$$

由定理 6.2 的推论可知指数 $L$ 遍历.

例 6.4 也是例 6.6 的特例.

**例 6.7**(演化算法(evolution algorithm))

假定我们有一个目标函数 $Q$(并不一定知道它的具体解析表达式,但是在自变量为任意 $x$ 时,我们通过实验可以测量到 $Q(x)$ 的值). 为求 $Q$ 的最小值可以先选取若干个(设为 $N$ 个)初值,排成一个 $N$ 维列向量,记为 $\xi$. 用随机迭代映射

$$\xi_{n+1} = f(\xi_n) + W_n,$$

$$f = (f^1,\cdots,f^N)^T,\ \xi_{n+1}^i = f^i(\xi_n) + W_n^i \quad (i = 1,2,\cdots,N),$$

其中一种最简单的做法是由某种优生原则从 $N$ 个分量中随机选出 $M$ 个分量来(如 $M=1$,或 $M=2$),余下分量中令 $f^i(x)=x$,$W_n^i=0$. 对选中的分量可令:$f(x)=\dfrac{Q(x+h)-Q(x)}{h}$,其中 $h$ 适当选取(若 $Q(x)$ 形式已知,则可用 $f(x)=\mathrm{grad}Q(x)$). 这种迭代方案来自生物遗传的思想:$x \to f(x)$ 称为从父代到子代的遗传,$W_n$ 称为变异. 从 $\xi_n \to \xi_{n+1}$ 称为演化. 从演化可以得到一个马氏链 $\xi$. 如果前面的条件满足,则 $\xi_n$ 有一极限分布 $F^{(\infty)}(y)$. 这还不能满足求 $Q$ 的最小值的需要. 为求 $\min Q(x)$,有多种方法,其中之一是模拟退火方法(前面已述其方法的大意),另一方法是加入竞争机制:令

$$\begin{cases} \eta_{n+1} = f(\xi_n) + W_n, & \text{称为变异;} \\ \xi_{n+1} = \begin{cases} \xi_n(x), & \text{若 } Q(\eta_{n+1}) \geqslant Q(\xi_n); \\ \eta_{n+1}, & \text{若 } Q(\eta_{n+1}) \leqslant Q(\xi_n) \text{ 称为遗传竞争.} \end{cases} \end{cases}$$

在 $M=1$,且 $f$ 满足一定条件下,用比较多一些的概率论知识,可以证明在这种竞争机制下:当 $n\to\infty$ 时 $\xi_n$ 按概率收敛到 $Q$ 取最小值的集合.

**例 6.8**(资金累积)

假定生产资金只在离散时刻进行统计.由于纯随机的原因,每个单位时刻的投资与产值的关系是有涨落的,而且假定它们是彼此独立同分布的.也就是:时刻 $n$ 投入资金 $x$ 于生产时,在时刻 $n+1$ 的产值为 $f(x,w_n)$,其中 $\{w_n\}$ 是只取有限个值 $1,2,\cdots,N$ 的独立同分布随机变量列,$P(w_n=k)=p_k>0(1\leqslant k\leqslant N)$.我们假定在时刻 $n+1$ 投入生产的全部资金(记为 $\xi_{n+1}(x)$)来自两部分:$n$ 时刻资金中的 $\theta$ 比例部分 $\theta\xi_n(x)(\theta<1)$;$n$ 到 $(n+1)$ 时刻产值 $f((1-\theta)\xi_n(x),w_n)$ 中的 $(1-\beta)$ 比例部分($\beta<1$,$\beta$ 比例部分用于消费,如工资、福利等等).即

$$\begin{cases}\xi_0(x)=x,\\ \xi_{n+1}(x)=\theta\xi_n(x)+(1-\beta)f((1-\theta)\xi_n(x),w_n).\end{cases}$$

下面我们来证明在以下条件下,这个资金积累过程是指数遍历的,也就是它以指数衰减的误差趋于一个分布稳定的情况,而达到随机的动态平衡.

假定 $f(x,k)$ 关于 $k$ 严格递增,而且在 $x\geqslant 0$ 时满足:

$$\begin{cases}f(0,k)=0,\\ \dfrac{\partial f}{\partial x}>0,\ \dfrac{\partial f}{\partial x}(\infty,k)=0,\\ \dfrac{\partial^2 f}{\partial x^2}<0,\\ \dfrac{\partial f}{\partial x}(0,k)>\dfrac{1}{1-\beta}.\end{cases}$$

设在初始时刻($n=0$)的资金为 $x$.为此只要验证定理 6.3 证明后面注的条件(6.17).记 $g_k(x)=\theta x+(1-\beta)f((1-\theta)x,k)$.那么 $g_k(0)=0$,且 $g_k$ 为非负凹函数($g_k''<0$).然而,由图形可以看出 $y=g_k(x)$ 与 $y=x$ 在 $0<x<\infty$ 有唯一的交点.记这个交点为 $a_k$,它满足 $g_k(a_k)=a_k$.又因为 $f(x,k)$ 关于 $k$ 是严格递增的,所以 $a_k$ 关于 $k$ 也是严格递增的.

利用 $g_k(a_k)=a_k$ 及 $g_k'(x)>0$ 易见：

$$x\geqslant a_1\Rightarrow\xi_n(x)\geqslant a_1,\quad x\leqslant a_N\Rightarrow\xi_n(x)\leqslant a_N.$$

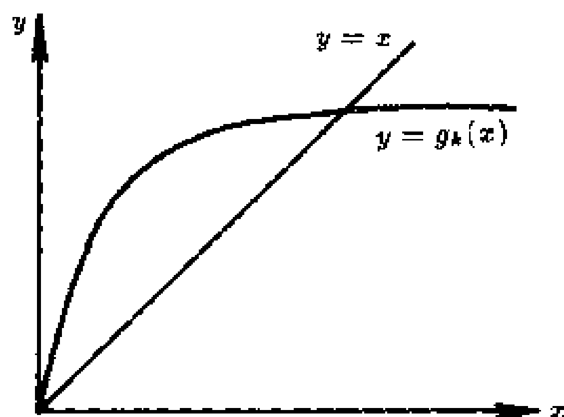



图 6.1

这样 $a_1\leqslant x\leqslant a_N\Rightarrow a_1\leqslant\xi_n(x)\leqslant a_N$. 从而可以认为马氏链 $\{\xi_n;n\geqslant 0\}$ 的取值区间为 $[a_1,a_N]$.

对于 $x\in[a_1,a_N]$，由图形看出 $g_k(x)\leqslant x$. 记

$$g_k^{(n)}(x)\triangleq g_k(g_k^{(n-1)}(x)),\quad g_k^{(1)}(x)=g_k(x).$$

由归纳法易见 $g_k^{(n)}(x)$ 对 $n$ 递减，从而有一个极限. 又由于上面的等式，这个极限必须是 $g_k$ 的不动点，即必须是 $a_k$，因此

$$g_k^{(n)}(x)\to a_k\quad(n\to\infty),\ \forall\ x\in[a_1,a_N].$$

特别地应该有

$$\begin{cases}g_1^{(n)}(a_N)\to a_1,\\ g_N^{(n)}(a_1)\to a_N,\end{cases}\quad n\to\infty.$$

于是 $\exists\ n_0$，使

$$g_1^{(n_0)}(a_N)<g_N^{(n_0)}(a_1).$$

并且对 $\forall\ z_0\in[g_1^{(n_0)}(a_N),g_N^{(n_0)}(a_1)]$ 有

$$\begin{aligned}P(\sup_{a_1\leqslant x\leqslant a_N}\xi_{n_0}(x)\leqslant z_0)&\geqslant P(\sup_{a_1\leqslant x\leqslant a_N}\xi_{n_0}(x)\leqslant g_1^{(n_0)}(a_N))\\&\geqslant P(w_1=1,\cdots,w_{n_0}=1)\\&=P(w_1=1)\cdots P(w_{n_0}=1)\\&=p_1^{n_0}.\end{aligned}$$

类似地

$$P(\inf_{a_1 \leqslant x \leqslant a_N} \xi_n(x) \geqslant z_0) \geqslant P(\inf_{a_1 \leqslant x \leqslant a_N} \xi_{n_0}(x) \geqslant g_N^{n_0}(a_1))$$

$$\geqslant P(w_1 = N, \cdots, w_{n_0} = N)$$

$$= p_N^{n_0}.$$

取 $\delta = p_N^{n_0} \wedge p_1^{n_0}$. 可知定理 6.3 后面的注中的(6.17)条件满足，因此以$[a_1, a_N]$为状态空间的马氏链$\{\xi_n : n \geqslant 0\}$为指数遍历的(记其极限分布函数为 $F^{(\infty)}$).

## §6.3 AR 模型(线性自回归)

下面讨论

$$\xi_{n+1} = \lambda\xi_n + W_n \quad (\{W_n\} \text{ 为 i.i.d. 随机变量序列}),$$

这时候在 §6.2 中的函数 $f(x) = \lambda x$ 是线性函数，因而不是有界的. 它不满足例 6.6 的条件. 对此我们有

**命题 6.4** 如果 $W_n$ 有界且有密度，$|\lambda| < 1$，那么存在分布函数 $F^{(\infty)}(y)$，使

$$\int_{-\infty}^{y} p^{(n)}(x, y)\mathrm{d}z \to F^{(\infty)}(y) \quad (x \to \infty, \forall\ x),$$

其中 $\xi_0 = x_0$.

**证明** 经过迭代我们有

$$\xi_n = \lambda^n x + \lambda^{n-1} W_1 + \cdots + \lambda W_{n-1} + W_n.$$

令 $\tilde{\xi}_n = \lambda^n x + W_1 + \lambda W_2 + \lambda^{n-1} W_n$. 那么 $\tilde{\xi}_n$ 与 $\xi_n$ 同分布. 由于 $W_n$ 们是有界的，$|\lambda| < 1$，所以

$$\tilde{\xi}_n \to \sum_{m=0}^{\infty} \lambda^m W_{m+1} \overset{\text{记为}}{=} \tilde{\xi}^{(\infty)} \quad (n \to \infty).$$

这里的收敛性这样理解：

$$E\left|\sum_{m=0}^{M} \lambda^m W_{m+1} - \sum_{m=0}^{N} \lambda^m W_{m+1}\right|^2 \quad (N < M)$$

$$= \sum_{m,n=N+1}^{M} \lambda^m \lambda^n E(W_{m+1} W_{n+1})$$

$$\leqslant \sum_{m,n=N+1}^{M} \lambda^m \lambda^n (EW_{m+1}^2 + W_{n+1}^2)/2$$

$$= \sum_{m=N+1}^{M} (\lambda^m)^2 EW_1^2 \to 0 \quad (\text{若 } N \to +\infty).$$

可见$\xi^{(\infty)}$是几乎处处有限，并且

$$I \triangleq E\left|\sum_{m=0}^{M} \lambda^m W_{m+1} - \xi^{(\infty)}\right|^2 \to 0 \quad (\text{若 } M \to +\infty).$$

再由 Chebechev 不等式有：

$$P\left(\left|\sum_{m=1}^{M} \lambda^m W_{m+1} - \xi^{(\infty)}\right| > \varepsilon\right) \leqslant \frac{I}{\varepsilon^2} \to 0 \quad (M \to \infty).$$

于是$\xi_n$的分布函数$\to \xi^{(\infty)}$的分布函数$F^{(\infty)}$. 从而$\xi_n$的分布函数$\int p^{(n)}(x,z)\mathrm{d}z \to F^{(\infty)}(y)$. 显见，由定义$F^{(\infty)}$与$x$无关.

**注 1**：关于这一段讨论的严格数学逻辑推导，读者可参考[YWL].

**注 2**：这个结果可以推广到多维. 设$\{W_n\}$是$R^d$中有界的独立同分布随机向量列，$x \in R^d$，

$$\begin{cases} \xi_{n+1} = A\xi_n + W_{n+1} & (A \text{ 为 } d \times d \text{ 矩阵}), \\ \xi_0 = x. \end{cases}$$

如果$A$的特征值的模均小于 1，那么命题 6.4 的结论也正确（证明方法类似）. 更易判别的是：设$\|A\| < 1$，其中$\|A\|$是矩阵$A$的欧氏模：$\|A\|^2 = \sum a_{ij}^2$. 若$\exists\, n_0$，使$\|A^{n_0}\| < 1$，则命题 6.4 的结论正确. $F^{(\infty)}$称为$\xi_n$的极限分布. 易证$\xi_n$是连续状态马氏链，并以$F^{(\infty)}$为其平稳初分布. 这方面的详细讨论还可见[BhW].

**说明**：注 1 中的条件$\alpha \triangleq \|A^{n_0}\| < 1$保证了

$$\|A^n\| \leqslant C\|A^{n_0}\|^{\left[\frac{n}{n_0}\right]} \quad (|\alpha| < 1).$$

从而保证了$\{\xi_n\}$的极限分布的存在. 这里$\|\cdot\|$有性质$\|AB\| \leqslant \|A\|\|B\|$可以由

$$\left(\sum_k a_{ik} b_{kj}\right)^2 \leqslant \left(\sum_k a_{ik}^2\right)\left(\sum_k b_{kj}^2\right)$$

得到.

**例 6.9**(一维 $p$ 阶 $AR$ 模型)(可参考第八章)

$$\xi_{n+1} = a_1\xi_n + a_2\xi_{n-1} + \cdots + a_p\xi_{n-p+1} + w_{n+1} \quad (n \geqslant p-1). \tag{6.22}$$

其中$\{W_n\}$为$R^1$中i.i.d.列且有界(有界性是这里为了简化而多加的条件),$\xi_0=x_0,\cdots,\xi_{p-1}=x_{p-1}$.

这个模型可以改写为向量形式,利用

$$\begin{aligned}
&\xi_{n+1} = a_1\xi_n + a_2\xi_{n-1} + \cdots + a_p\xi_{n-p+1} + W_{n+1},\\
&\xi_n = \xi_n,\\
&\cdots\cdots\\
&\xi_{n-p+2} = \xi_{n-p+2}.
\end{aligned} \tag{6.23}$$

记

$$\eta_n \triangleq \begin{pmatrix} \xi_n \\ \xi_{n-1} \\ \vdots \\ \xi_{n-p+1} \end{pmatrix},$$

则上式成为 $k$ 维向量方程.

$$\eta_{n+1} = \begin{pmatrix} a_1 & \cdots\cdots & a_k \\ 1 & 0 & \\ & \ddots & \ddots \\ & 1 & 0 \end{pmatrix} \eta_n + \begin{pmatrix} W_n \\ 0 \\ \vdots \\ 0 \end{pmatrix} \triangleq A\eta_n + \epsilon_{n+1}.$$

这里的矩阵 $A$ 的形式称为自然典则形式,它的特征多项式为下述行列式

$$\det(A - \lambda I) = (-1)^p\lambda^p - a_1\lambda^{p-1} - \cdots - a_{p-1}\lambda - a_p.$$

所以只要 $\lambda^p - a_1\lambda^{p-1} - \cdots - a_p = 0$ 的根均在$|\lambda|<1$中,那么$\eta_n$有不依赖$x_0,\cdots,x_{p-1}$极限分布,我们把它记为$F^{(p)}$.因而其分量$\xi_n$也有不依赖$x_0,\cdots,x_{p-1}$的极限分布,我们把它记为$F^{(1)}$,即$F^{(1)}(y)\triangleq F^{(p)}(y,+\infty,\cdots,+\infty)$.注意$\eta_n$是连续状态马氏链,然

而 $\xi_n$ 不是马氏链，但是它是一个 $p$ 维马氏链 $\eta_n$ 的一个分量，称为 $p$ 步马氏链.

**注 1**：利用 $A$ 的 Jordan 标准形，不难证明若 $A$ 的特征根的模均小于 1 时，一定存在 $n_0$，使 $\|A^{n_0}\|<1$.

**注 2**：稍用一点严格的概率论的知识(Borel-Cantelli 引理及强大数律)可以证明在命题 6.4(多维情形)中，若将条件"$W_n$ 有界"用"$\exists\ a>0$ 使 $E|w_n|^a<\infty$"代替，则结论仍然正确.

AR 模型可以很好地用来对于语音信号所代表的语义作人工识别. 有人认为表明语音信号的语言信息主要包含在系数参数 $(a_1,a_2,\cdots,a_p)$ 中，而"残差"$W_n$ 则包含了更多的说话人的信息，如果不仅要知道声信号代表的语言含义，而且还要知道是哪一个人说的，下一节(§6.4)中的 ARMA 模型提供了一个有价值的途径. 但也有实践表明语音信息与说话人的信息都包含在足够高阶($p$ 大)的 AR 系数中. 如何选取 $p$，这是应用时间序列广泛讨论的一个问题，有许多不同的统计方法. 确定了 $p$ 如何估 $(a_1,a_2,\cdots,a_p)$? 我们将在第八章中作简要的介绍.

对 AR 模型、ARMA 模型的进一步学习，读者可以参考时间序列方面的有关著作，如[X]，[Fu]等.

## §6.4 ARMA 模型

ARMA$(p,q)$是一个非常广泛应用的近似模型，它的数学表达为：

$$\begin{aligned}\xi_{n+p}=a_1\xi_{n+p-1}+\cdots+a_p\xi_n+W_{n+p}\\+b_1W_{n+p-1}+\cdots+b_qW_{n+p-q},\end{aligned}\tag{6.24}$$

其中 $n\geqslant q-p$，$\{W_n\}$ 是 i.i.d. 随机变量列，而且与随机向量 $(\xi_1,\cdots,\xi_{p-1})$ 相互独立. 这里为了避开技术性的困难，我们还假定 $\{W_n\}$ 是有界随机变量列.

我们把(6.24)改写为($n\geqslant0$)：

$$
(\xi_{n-p},\cdots,\xi_{n-1},W_{n-p},W_{n+p+1},\cdots,W_{n-p-q+1})^T
$$

$$
=\begin{pmatrix}
a_1 & \cdots & a_{p-1} & a_p & b_1 & \cdots & b_{q-1} & b_q \\
1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0
\end{pmatrix}
\begin{pmatrix}
\xi_{n-p+1} \\ \vdots \\ \xi_n \\ W_{n-p+1} \\ W_{n-p-2} \\ \vdots \\ W_{n-p-q}
\end{pmatrix}
+
\begin{pmatrix}
W_{n+p} \\ 0 \\ \vdots \\ 0 \\ W_{n-p} \\ 0 \\ \vdots \\ 0
\end{pmatrix}
\tag{6.25}
$$

并且它简单地记成

$$
\eta_{n+1}=\begin{pmatrix} A & B \\ O & C \end{pmatrix}\eta_n+\varepsilon_{n+1}, \tag{6.25}
$$

其中 $O$ 代表元素全零的矩阵

$$
A=\begin{pmatrix}
a_1 & a_2 & \cdots & a_{p-1} & a_p \\
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & 1 & 0
\end{pmatrix},\quad
B=\begin{pmatrix}
b_1 & \cdots\cdots & b_q \\
0 & \cdots\cdots & 0 \\
\cdots & \cdots & \cdots \\
0 & \cdots\cdots & 0
\end{pmatrix},
$$

$$
C=\begin{pmatrix}
0 & 0 & \cdots & 0 & 0 \\
1 & 0 & \cdots & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & 1 & 0
\end{pmatrix}.
$$

$$
\varepsilon_{n+1}=(W_{n+1},0,\cdots,0,W_{n+1},0,\cdots,0)^T.
$$

这就纳入了(6.23)的框架. 注意

$$
\det\left(\begin{pmatrix} A & B \\ O & C \end{pmatrix}-\lambda\begin{pmatrix} I & \\ & I \end{pmatrix}\right)=\det(A-\lambda I)(-\lambda)^q
$$

$$
=(\lambda^p-a_1\lambda^{p-1}-\cdots-a_p)(-\lambda)^q.
$$

所以只要 $\lambda^p-a_1\lambda^{p-1}-\cdots-a_p=0$ 的根 $\lambda$ 满足 $|\lambda|<1$,那么 $\eta_n$ 就有不依赖初值的极限分布,我们把它记成

$$F^{(\infty)}(y_1,\cdots,y_p,v_1,\cdots,v_q).$$

从而我们知道 $\xi_n$ 有一个极限分布 $F^{(\infty)}(y)$，而且

$$F^{(\infty)}(y)=F^{(\infty)}(y,+\infty,\cdots,+\infty).$$

本章的讨论大多涉及马氏链的指数遍历性及指数 $L^1$ 遍历性，这种收敛性要求并不太强. 在实际问题中有时要用到更强的收敛性（如概率收敛，概率为 1 的收敛），这就需要更多的概率论知识.

## 习　题

1. 设 $\{W_n\}$ 为 i.i.d. 列，具有正连续密度，又 $\xi_{n+1}=f(\xi_n)+W_n \pmod 1$，则连续状态马氏链 $\xi_n$ 为指数 $L^1$ 遍历的.

2. 设 $\{W_n\}$ 为正 i.i.d. 列，具有连续密度 $\rho$，又若有 $0<a\leqslant f\leqslant b$，$\xi_{n+1}=f(\xi_n)W_n$，则连续状态马氏链 $\{\xi_n:n\geqslant 0\}$ 为指数 $L^1$ 遍历的.

3. 设 $\{W_n\}$ 为 i.i.d. 列，具有正连续密度 $\rho$，$f'(x)>0$. $g$ 有界，$f(\xi_{n+1})=g(\xi_n)+W_n$，则连续状态马氏链 $\xi_n$ 是指数 $L^1$ 遍历的.

4. $W_n$ 如前，记其分布密度为 $\rho(y)$. $\dfrac{\partial f}{\partial y}>0$，又若当 $x$ 固定后设 $f(x,y)$ 的反函数为 $h(x,y)$，且满足 $\exists\ g(y)\geqslant 0$ 使 $\rho(h(x,y))\geqslant \dfrac{\partial f}{\partial y}g(y)$. 则马氏链 $\xi_n$ 指数 $L^1$ 遍历，其中 $\xi_n$ 由下式决定：

$$\xi_{n+1}=f(\xi_n,w_n).$$

5. 设 $W_n$ 为 i.i.d. 列，$W_n\sim\begin{pmatrix}-1 & 1\\ \dfrac{1}{2} & \dfrac{1}{2}\end{pmatrix}$. 令

$$f(x)=\begin{cases}x+1, & x\in[-2,0];\\ x-1, & x\in[0,2],\end{cases}$$

$$\begin{cases}\xi_{n+1}=f(\xi_n)+w_n.\\ \xi_0=x\end{cases}$$

的解 $\xi_n$ 记为 $\xi_n(x)$，证明：$\{\xi_n(x):n\geqslant1\}$ 为 i.i.d.，且当 $x\in(0,2]$ 时，

$$\xi_n(x)\sim\pi_x\triangleq\begin{pmatrix}x-2 & x\\ 0.5 & 0.5\end{pmatrix};$$

而当 $x\in[-2,0)$ 时，

$$\xi_n(x)\sim\pi_{x+2}\left(=\begin{pmatrix}x & x+2\\ 0.5 & 0.5\end{pmatrix}\right).$$

6. 在上题中，若 $w_n$ 遵从[-1,1]均匀分布，证明对于 $x\in[-2,2]$，$\{\xi_{2n}(x):n\geqslant1\}$ 为 i.i.d.，具分布密度（不依赖 $x$）

$$\rho(y)=(2-|y|)/4\quad(-2\leqslant y\leqslant2).$$

7. 设 $W_n,\pi_x$ 及 $f$ 如题 5. 但是函数 $f_\delta$ 定义如下：

$$f_\delta(x)=\begin{cases}f(x), & x\in[-2,2]\setminus(-\delta,\delta);\\ -x, & x\in(-\delta,\delta).\end{cases}$$

相应于 $f_\delta$ 的 $\xi_n$ 记为 $\xi_n^\delta$. 证明：当 $x\in[-1,1]$ 时，$\{\xi_n^\delta(x):n\geqslant1\}$ 为 i.i.d.，且

$$\xi_n^\delta(x)\sim\begin{cases}\pi_{x+2}, & x\in[-1,-\delta];\\ \pi_{1-x}, & x\in(-\delta,\delta);\\ \pi_x, & x\in[\delta,1],\end{cases}$$

而当 $x\in[-2,2]\setminus[-1,1]$ 时，$\{\xi_n^\delta(x):n\geqslant2\}$ 为 i.i.d.，且

$$\xi_n^\delta(x)\sim\begin{cases}\pi_{x+2}, & x\in[-2,-1);\\ \pi_x, & x\in(1,2].\end{cases}$$

8. 证明例 6.9 中的 $\eta_n$ 是连续状态马氏链.

9. 以例 6.9 中的 $\xi_n$ 为模本，给出 $k$ 步（连续状态）马氏链的概念.

10. 设 $\{W_n\}$ 为 i.i.d.，具有分布密度 $\rho(y)$，

$$\xi_{n+k}=f(\xi_n,\cdots,\xi_{n+k-1})+W_n,$$

$$\xi_0=x_0,\cdots,\xi_{k-1}=x_{k-1},$$

求证 $\{\xi_n:n\geqslant0\}$ 为 $k$ 步马氏链.

11. 设 $A=(a_{ij})$，$B=(b_{ij})$ 为 $n$ 阶方阵. 令矩阵的模的平方

$\|\cdot\|^2$为其全体元素的平方和,试证明:

(1) $\|AB\| \leqslant \|A\|\,\|B\|$.

(2) 对任意的 $n$ 维向量 $x$,令 $|x|^2 = \sum\limits_{i=1}^{n} x_i^2$,则

$$\|A\| = \sup_{|x|=1} |Ax|.$$

(3) $\|A\| = \alpha < 1$, $\left|\sum\limits_{n=0}^{\infty} A^n x\right| \leqslant \dfrac{|x|}{1-\alpha}$.

# 第七章　平稳序列、保测映射与遍历论初步

从历史上看，本章内容源于平稳态统计物理与平稳过程(序列)的研究．但是，近年来从动力体系的研究已发展出一个独立的重要学科——遍历论．它在非线性科学等领域的研究中起了重要作用．大部分随机过程的教材不涉及这部分内容．但笔者以为，平稳马氏链是这部分内容的一个极好特例，它具有典型性、对于理解与认识一般概念与理论很有启发性，再则随机过程与遍历论两者的思想方法可以相辅相成．看来，它们在更高一级水平上的结合是近年来，并将是今后若干年内的一个新动向．为此，本书中我们设法将这个领域中的基本概念与部分基本思想在本章作一个简要的讨论．

## §7.1　平稳序列与保测映射

让我们从平稳马氏链这个特例出发．

**例 7.1(平稳马氏链)**　设 $\zeta=\{\zeta_n;n=0,1,\cdots\}$ 是一个 $N$ 状态的时齐马氏链，$P=(p_{ij})$ 是它的一步转移阵．由 §3.3 定理 3.13 可见：一定存在不变概率分布 $\underset{\vee}{\pi}=(\pi_1,\cdots,\pi_N)$，使得(3.19)

$$\sum_{k=1}^{N}\pi_k p_{kj}=\pi_j\quad(1\leqslant j\leqslant N)$$

成立．再由命题 3.10，就可知道，若以 $\underset{\vee}{\pi}$ 为 $\zeta$ 的初分布，即

$$P(\zeta_0(w)=i)=\pi_i\quad(1\leqslant i\leqslant N).$$

则对任意的 $k\geqslant1,s\geqslant1$，及时刻 $n_1,\cdots,n_k$，$\{\zeta_{n_1}(w),\cdots,\zeta_{n_k}(w)\}$ 与 $\{\zeta_{n_1+s}(\omega),\cdots,\zeta_{n_k+s}(\omega)\}$ 都同分布．从而，$\zeta$ 与 $\tilde{\zeta}\triangleq\{\zeta_{n+1}(\omega),n=0,1,$

$\cdots\}$同分布，也就是 $\zeta$ 的分布对时间推移不变．这种对时间推移分布不变的性质叫做**平稳性**．下面我们给出一般定义．

**定义 7.1（平稳序列）** 随机变量列 $\{\zeta_n(\omega); n \geqslant N_0\}$ 如果满足下述条件则称为一个平稳序列：对任意的整数 $k \geqslant 1$ 与 $s \geqslant 1$ 及时刻 $n_1, \cdots, n_k$，$\{\zeta_{n_1}(\omega), \cdots, \zeta_{n_k}(\omega)\}$ 与 $\{\zeta_{n_1+s}(\omega), \cdots, \zeta_{n_k+s}(\omega)\}$ 都同分布．

**定义 7.1′（平稳过程）** 随机过程 $\{\zeta_t(\omega); t \geqslant T_0\}(T_0 > -\infty)$ 称为平稳的如果对任意整数 $k \geqslant 1$ 与 $s > T_0$ 及时刻 $n_1, \cdots, n_k$，$\{\zeta_{n_1}(\omega), \cdots, \zeta_{n_k}(\omega)\}$ 与 $\{\zeta_{n_1+s}(\omega), \cdots, \zeta_{n_k+s}(\omega)\}$ 都同分布．

**例 7.2** 若 $\zeta$ 是例 1 中的平稳马氏链，$f$ 与 $f_k(k=1,\cdots,r)$ 是实函数．则 $\{f(\zeta_n(\omega)); n=0,1,\cdots\}$ 与 $\{(f_1(\zeta_n(\omega)), \cdots, (f_r(\zeta_n(\omega))); n=0,1,\cdots\}$ 分别是平稳的实值与 $r$-维向量值随机序列．

**例 7.3**（马氏链生成的符号动力体系及对应的保测映射）

对于例 1 中的平稳马氏链 $\zeta$，令 $S \triangleq \{1,2,\cdots,N\}$，$\Omega = \{(\omega_0, \cdots, \omega_n, \cdots); \omega_n \in S; n=0,1,2,\cdots\}$ 对于任意整数 $m \geqslant 1$，$a_0, \cdots, a_m \in S$，我们再令

$$A_{a_0,\cdots,a_m} = \{(\omega_0, \cdots, \omega_n, \cdots); \omega_k = a_k, k = 0, \cdots, m\},$$

且对于 $A_{a_0,\cdots,a_m}$ 定义概率：

$$P(A_{a_0,\cdots,a_m}) = \pi_0 p_{a_0 a_1} p_{a_1 a_2} \cdots p_{a_{m-1} a_m}.$$

于是类似于第一章 §1.3，我们可以有一个事件体 $\mathscr{F}$，及 $\mathscr{F}$ 上定义的概率 $P$，即有概率空间 $(\Omega, \mathscr{F}, P)$．令：$\boldsymbol{A} = (a_{ij})$ 是一个 $N$ 阶方阵，其中

$$a_{ij} = \begin{cases} 1, & P(\xi_{n+1} = j | \xi_n = i) > 0; \\ 0, & P(\xi_{n+1} = j | \xi_n = i) = 0. \end{cases}$$

令

$$\bar{\Omega} = \{(\omega_0, \omega_1, \cdots, \omega_n, \cdots); \quad a_{\omega_n \omega_{n+1}} > 0, \forall n \in Z\}.$$

容易证明 $P(\bar{\Omega}) = 1$．令 $\bar{\mathscr{F}} = \{A \cap \bar{\Omega}; A \in \mathscr{F}\}$．于是我们可以得到

概率空间$(\Omega,\mathscr{F},P)$.

令$\zeta_n(\omega)=\omega_n$，则$\{\zeta_n(\omega)\}$就是一个例 7.1 中那样在$(\Omega,\mathscr{F},P)$上的平稳马氏链. 再定义$\Omega$上映射$\theta$：

$$\theta:\ \Omega\to\Omega,\ \omega=(\omega_0,\cdots,\omega_n,\cdots)\to\theta\omega\triangleq(\omega_1,\cdots,\omega_{n+1},\cdots).$$

我们考察下列集合：

$$\begin{aligned}\theta^{-1}A_{a_0,\cdots,a_m}&=\{\theta\omega\in A_{a_0,\cdots,a_m}\}\\&=\{\omega=(\omega_0,\cdots,\omega_n,\cdots):(\omega_1,\cdots,\omega_{n+1},\cdots)\in A_{a_0,\cdots,a_m}\}\\&=\{\omega:\zeta_1(\omega)=\omega_1=a_0,\cdots,\zeta_{m+1}(\omega)=\omega_{m+1}=a_m\}\\&=\{\omega:\zeta_1(\omega)=a_0,\cdots,\zeta_{m+1}=a_m\}.\end{aligned}$$

由$\zeta$的平稳性就知道：它的概率

$$\begin{aligned}P(\theta^{-1}A_{a_0,\cdots,a_m})&=P(\{\omega:\zeta_1(\omega)=a_0,\cdots,\zeta_{m+1}=a_m\})\\&=P(\{\omega:\zeta_0(\omega)=a_0,\cdots,\zeta_m(\omega)=a_m\})\\&=P(\{\omega:\omega_0=a_0,\cdots,\omega_m=a_m\})\\&=P(A_{a_0,\cdots,a_m}).\end{aligned}$$

上式说明在映射$\theta$下，集合$A_{a_0,\cdots,a_m}$的原象集的概率与$A_{a_0,\cdots,a_m}$的概率相同. 可以证明对$\mathscr{F}$中的任何集合$A$都有性质

$$P(\theta^{-1}A)=P(A). \tag{7.1}$$

也就是在映射$\theta$下，概率（也称测度）$P$是保持不变的. $\Omega=\{(\omega_0,\omega_1,\cdots,\omega_n,\cdots):\omega_k\in S;k=0,1,2,\cdots\}\triangleq S^{\mathbb{Z}_+}$与$\Omega$上的映射$\theta$称为一个**符号动力体系**. $\theta$也称为**推移映射**.

**定义 7.2（保测映射）** 设有概率空间$(\Omega,\mathscr{F},P)$及$\Omega\to\Omega$的映射$\theta$，使得对任何$A\in\mathscr{F}$，都有$\theta^{-1}A\in\mathscr{F}$，而且保持测度相等：$P(\theta^{-1}A)=P(A)$，则称$\theta$是一个$(\Omega,\mathscr{F},P)$上的保测映射.

例 7.3 中的推移映射$\theta$显见是保测映射的一个典型例子.

**命题 7.1** 设$\theta$是$(\Omega,\mathscr{F},P)$上的一个保测映射，$\zeta(\omega)$是$(\Omega,\mathscr{F},P)$上的任意一个随机变量. 令$\xi_0(\omega)\triangleq\xi(\omega)$，$\xi_1(\omega)\triangleq\xi(\theta\omega)$，

$\cdots, \xi_n(\omega) \triangleq \xi_{n-1}(\theta\omega) = \xi(\theta^n\omega), \cdots$,则 $\{\xi_n(\omega); n=0,1,\cdots\}$ 是一个平稳序列.

**证明** 考查

$$
\begin{aligned}
\tilde{A} &\triangleq \{\omega: \xi_1(\omega) \leqslant x_1, \cdots, \xi_n(\omega) \leqslant x_n\} \\
&= \{\omega: \xi_0(\theta\omega) \leqslant x_1, \cdots, \xi_{n-1}(\theta\omega) \leqslant x_n\} \\
&= \theta^{-1}\{\omega: \xi_0(\omega) \leqslant x_1, \cdots, \xi_{n-1}(\omega) \leqslant x_n\} \\
&\triangleq \theta^{-1}A.
\end{aligned}
$$

显然 $A \in \mathscr{F}$,由 $\theta$ 的保测性假定,我们有

$$
P(A) = P(\theta^{-1}A) = P(\tilde{A}).
$$

由归纳法易见,

$$
\begin{aligned}
P(A) &= P(\theta^{-1}A) = P(\theta^{-1}(\theta^{-1}A)) \\
&= P(\{\omega: \zeta_1(\theta\omega) \leqslant x_1, \cdots, \zeta_n(\theta\omega) \leqslant x_n\}) \\
&= P(\{\omega: \zeta_2(\omega) \leqslant x_1, \cdots, \zeta_{n+1}(\omega) \leqslant x_n\}) = \cdots \\
&= P(\{\omega: \zeta_{s-1}(\theta\omega) \leqslant x_1, \cdots, \zeta_{n-2+s}(\theta\omega) \leqslant x_n\}) \\
&= P(\{\omega: \zeta_s(\omega) \leqslant x_1, \cdots, \zeta_{n-1+s}(\omega) \leqslant x_n\}).
\end{aligned}
$$

可见 $(\zeta_0(\omega), \cdots, \zeta_{n-1}(\omega))$ 与 $(\zeta_s(\omega), \cdots, \zeta_{n-1+s}(\omega))$ 是同分布的. 又由于 $(\zeta_{n_1}(\omega), \cdots, \zeta_{n_k}(\omega))$ 的分布只不过是 $(\zeta_0(\omega), \zeta_1(\omega), \cdots, \zeta_{n_k}(\omega))$ 的边缘分布,从而 $(\zeta_{n_1}(\omega), \cdots, \zeta_{n_k}(\omega))$ 与 $(\zeta_{n_1+s}(\omega), \cdots, \zeta_{n_k+s}(\omega))$ 也全是同分布的,即 $\zeta$ 是一个平稳序列.

上面这个命题说明 $(\Omega, \mathscr{F}, P)$ 上一个保测映射 $\theta$ 与 $(\Omega, \mathscr{F}, P)$ 上的任意一个随机变量 $\zeta_0(w)$ 都可以生成一个以 $\zeta_0(w)$ 为初值的平稳序列. 相反地,事实上对任何一个 $(\Omega, \mathscr{F}, P)$ 上的平稳序列 $\{\zeta_n(\omega)\}$,都可以构造一个概率空间 $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$,及其上面一个保测映射 $\theta$ 和与 $\zeta_0(\omega)$ 同分布的一个随机变量 $\tilde{\zeta}_0(\omega)$,而且只要令

$$
\tilde{\zeta}_n(\omega) = \tilde{\zeta}_0(\theta^n\omega) = \tilde{\zeta}_{n-1}(\theta\omega), \tag{7.2}
$$

就有 $\{\tilde{\zeta}_n(\omega); n=0,1,2,\cdots\}$ 与 $\{\zeta_n(\omega); n=0,1,2,\cdots\}$ 同分布. 在这

个意义上，我们可以认为平稳序列都是由保测映射生成的，也就是说平稳序列与保测映射在实际上是互相对应的.

## §7.2 Birkhoff 遍历定理

Birkhoff 遍历定理是保测映射和平稳序列最深刻的基本性质之一，它具有广泛的应用. 在本节中，我们将从分析平稳马氏链这一特殊情况下的相应结果出发，进而阐述一般保测映射与平稳序列的 Birkhoff 遍历定理及其意义. 由于此定理的证明涉及较为深入的测度论工具，它已远远超出本书的宗旨与水平，我们只能略去它了. 有兴趣的读者可参考有关书藉(例如，笔者所写的《随机过程引论》一书).

### 1. 有限状态平稳马氏链的 Birkhoff 定理

考虑 §7.1 中例 7.1 与例 7.3 中的有限状态马氏链及相应的保测映射，即考虑概率空间$(\Omega,\mathscr{F},P)$及其上的 $N$ 状态马氏链 $\zeta=\{\zeta_n(\omega)\}$，其中记

$$S=\{1,2,\cdots,N\},$$

$$\Omega=\{\omega=(\omega_0,\cdots,\omega_n,\cdots);\omega_k\in S;k=0,1,2,\cdots\},$$

$$\zeta_n(\omega)=\omega_n,$$

$$\theta:\ \Omega\to\Omega,\theta(\omega_0,\omega_1,\cdots,\omega_n,\cdots)=(\omega_1,\omega_2,\cdots,\omega_{n+1},\cdots),$$

$$P(\zeta_{n+1}(\omega)=j|\zeta_n(\omega)=i)=p_{ij},$$

$$\pi_i=P(\zeta_n(\omega)=i).$$

由 §3.4 中的结果我们可以知道：对于 $\zeta(\omega)$的常返状态 $i$ 应该有.

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}f(\zeta_k(\omega))=\sum_{j\in H}\frac{f(j)}{\mu_j}. \tag{7.3}$$

而其中 $H$ 是 $i$ 所在的常返类，$\mu_i=L_{ii}^{-1}$，而 $L_{ii}=\lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{k=1}^{n}P_{ii}(k)$.

我们还知道，当$\{\zeta_n(\omega)\}$是一个互通的有限状态马氏链时，对于任

意的 $i,j\leqslant N$，有 $L_{ij}=L_{oj}$. 而且 $\underset{\sim}{\pi}\triangleq(L_{11},\cdots,L_{NN})$ 是 $\zeta=\{\zeta_n\}$ 唯一的不变概率分布. 从而对于只取有限个值的平稳不可约马氏链就有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}f(\zeta_k(\omega))=\sum_{j\in H}\frac{f(j)}{\mu_j}=\sum_{k=1}^{n}f(i)L_{ii}$$

$$=\sum_{k=1}^{n}f(i)\pi_i=Ef(\zeta_n(\omega)).\qquad(7.4)$$

如果 $\{\zeta_n\}$ 具有若干不可约常返类 $H_1,\cdots,H_s$，由定理 3.15 后的说明我们可以看出

$$L_{ij}=\begin{cases}\pi_j^{(k)}, & 当\ i,j\in H_k;\\ 0, & 当\ i\ 常返且\ i,j\ 在不同一个常返类时.\end{cases}$$

另一方面，若 $\underset{\sim}{\pi}$ 是 $(\pi_1,\cdots,\pi_N)$ 是 $\{\zeta_n\}$ 的平稳初分布，则一定有 $\sum_{i=1}^{n}p_{ij}\pi_i=\pi_j$，而且由归纳法易得到：对任意的 $n\geqslant1$，

$$\sum_{i=1}^{n}p_{ij}(n)\pi_i=\sum_{i=1}^{n}\pi_i\sum_{k=1}^{n}p_{ik}p_{kj}(n-1)\cdots=\sum_{k=1}^{n}p_{kj}\pi_k=\pi_j.$$

再利用定理 3.12，对于常返且 $i,j\in H_k$ 应有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{n}p_{ij}(l)=L_{ij}\triangleq\pi_j^{(k)}.$$

由此就得到

$$\pi_j=\frac{1}{n}\sum_{i=1}^{N}\pi_i\sum_{l=1}^{n}p_{ij}(l)\to\sum_{i\in H_k}\pi_i\pi_j^{(k)}=\pi_j^{(k)}\sum_{i\in H_k}\pi_i\triangleq\pi_j^{(k)}\alpha_k.$$

$$(7.5)$$

其中 $\alpha_k=\sum_{i\in H_k}\pi_i$ 将 $\pi_j^{(k)}$ 扩大定义为：

$$\bar{\pi}_j^{(k)}=\begin{cases}\pi_j^{(k)}, & j\in H_k;\\ 0, & j\overline{\in}H_k.\end{cases}$$

令 $\underset{\sim}{\pi}^{(k)}\triangleq(\bar{\pi}_1^{(k)},\cdots,\bar{\pi}_N^{(k)})$，于是 $\underset{\sim}{\pi}=\sum_{k=1}^{s}\alpha_k\underset{\sim}{\pi}^{(k)}$. 因为 $\sum_k\alpha_k=1$，可见任何 $\{\zeta_n\}$ 的不变分布都是 $\underset{\sim}{\pi}^{(1)},\cdots,\underset{\sim}{\pi}^{(s)}$ 的凸线性组合. 全体不变概

率分布的集合就是$\pi^{(1)},\cdots,\pi^{(s)}$的凸线性闭包(对于可数状态的马氏链,这也意味着对于非常返状态 $i$,有 $\pi_i=0$,从而可以看出一个平稳马氏链,从非常返状态出发的概率必须是 0).这时(7.3)式就意味着从不同的常返类出发(7.3)右端的极限是不同的:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{n} f(\zeta_l(\omega)) = \sum_{j\in H_k}\pi_j^{(k)}f(j) \quad (\zeta_0(\omega)\in H_k). \qquad (7.6)$$

对于时齐马氏链,下列事件

$$\Omega_k=\{\omega:\zeta_0(\omega)\in H_k\},\ \Omega_k^{(n)}=\{\omega:\zeta_n(\omega)\in H_k,\forall\, n\geqslant 0\}$$

之间只差零测集,可见 $\Omega_k$ 在映射 $\theta$ 下是概率为 1 地不变的,即

$$P((\theta^{-1}\Omega_k\setminus\Omega_k)\cup(\Omega_k\setminus\theta^{-1}\Omega_k))=0. \qquad (7.7)$$

以后简称概率为 1 地不变的集合为**不变集**.于是(7.5)可改写为,以概率为 1 地成立下式:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{n} f(\zeta_l(\omega)) = \sum_{k=1}^{s} I_{\Omega_k}(\omega)\sum_{j}\pi_j^{(k)}f(j). \qquad (7.8)$$

将(7.8)式右端记为$g(\omega)$,那么它应以概率为 1 地满足

$$g(\theta\omega) = \sum_{k=1}^{s} I_{\Omega_k}(\theta\omega)\sum_{j}\pi_j^{(k)}f(j) = \sum_{k=1}^{s} I_{\theta^{-1}\Omega_k}(\omega)\sum_{j}\pi_j^{(k)}f(j)$$

$$= \sum_{k=1}^{s} I_{\Omega_k}(\omega)\sum_{j}\pi_j^{(k)}f(j) = g(\omega). \qquad (7.9)$$

如果将 $f(\zeta_0(\omega))$记为 $g(\omega)$,那么

$$f(\zeta_n(\omega)) = f(\zeta_0(\theta^n\omega)) = g(\theta^n\omega). \qquad (7.10)$$

于是(7.8)与(7.9)就可用 $\theta$ 表示为:以概率为 1 地有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} g(\theta^k\omega) = \bar{g}(\omega), \qquad (7.11)$$

$$\bar{g}(\theta\omega) = \bar{g}(\omega). \qquad (7.12)$$

在(7.11)两端分别求数学期望就得到

$$E\bar{g}(\omega) = \lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{n} Eg(\theta^l\omega)$$

$$= E\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} g(\theta^k\omega)\right) = E\bar{g}(\omega). \qquad (7.13)$$

(7.13)中数学期望与极限运算可以交换的原因是 $g(\theta^n\omega)=f(\zeta_0(\theta^n\omega))$，一共只有有限个不同的取值，从而

$$Eg(\theta^n\omega) = \sum_{l=1}^{n} f(l)P(\zeta_n(\omega)=l)$$

无非是一个有限项线性组合运算.它当然可以与有限运算交换次序(应该指出的是：一般来讲，不能随意将数学期望与极限运算交换次序).

下面我们再来看随机变量 $\bar{g}(\omega)$在什么条件下能像(7.4)中那样以概率为 1 地不依赖于初始状态 $\zeta_0(\omega)$，而是简单地等于$Eg(\omega)$(由 $\zeta_n$ 的平稳性，对于任意 $n\geqslant 0$，它应该等于 $Ef(\zeta_n(\omega))$). 仔细考查(7.8)式就可发现

$$\begin{aligned}
\bar{g}(\omega) &= \sum_{k=1}^{r} I_{\Omega_k}(\omega) \sum_j \pi_j^{(k)} f(j) \\
&= \sum_{k=1}^{r} I_{\Omega_k}(\omega) \sum_j \frac{\pi_j}{\sum_{i\in H_k}\pi_i} f(j) \quad (\text{见(7.5)式}) \\
&= \sum_{k=1}^{r} I_{\Omega_k}(\omega) \sum_j P(\zeta_0(\omega)=j\in H_k \mid \zeta_0(\omega)\in H_k) f(j) \\
&= \sum_{k=1}^{r} I_{\Omega_k}(\omega) \sum_j P(\zeta_0(\omega)=j \mid \Omega_k) f(j). \qquad (7.14)
\end{aligned}$$

也就是说$\bar{g}(\omega)$在每一个 $\Omega_k$ 上，其实是 $f(\zeta_0(\omega))=g(\omega)$在 $\Omega_k \triangleq \{\zeta_0(\omega)\in H_k\}$中的条件数学期望. $\bar{g}(\omega)$概率为 1 地不依赖初状态 $\zeta_0(\omega)$，当且仅当有一个 $k_0$，使 $\sum_{j\in H_{k_0}}\pi_j = P(\zeta_0(\omega)=j\in H_{k_0})=1$，也就是$\zeta_0(\omega)$以概率为 1 地集中在一个不可约常返类中. 这时不变集 $\Omega_{k_0}$的概率

$$P(\Omega_{k_0}) = \sum_{j\in H_{k_0}} \pi_j = \begin{cases} 1, & k = k_0; \\ 0, & \text{其他}. \end{cases}$$

也就是不变集合只可能是概率为 0 或 1 的集合. 反之, 若 $P(\Omega_k) = \sum_{j\in H_k} \pi_j > 0$ 对两个或两个以上的 $k$ 成立, 那么 $g(\omega)$ 一定依赖初状态 $\zeta_0(\omega)$. 进而我们还可以看到 (7.11) 中的 $g(\omega)$ 并不一定必须是 $f(\zeta_n(\omega))$ 这种形式的. 若

$$g(\omega) = f(\zeta_0(\omega), \zeta_1(\omega), \cdots, \zeta_n(\omega)),$$

(7.11) 仍然成立. 要证明这一点, 只要令

$$\eta_n(\omega) = (\zeta_n(\omega), \cdots, \zeta_{n+m}(\omega)),$$

容易看出 $\eta_n(\omega)$ 仍然只能取有限种不同状态 ($N^{m+1}$ 种), 而且

$$\begin{aligned} P(\eta_{n+1}(\omega) &= (j_0, \cdots, j_m) \mid \eta_n(\omega) = (i_0, \cdots, i_m), \eta_{n-1}, \cdots) \\ &= P(\zeta_{n+1}(\omega) = j_0, \cdots, \zeta_{n+m+1}(\omega) = j_m \mid \zeta_n(\omega)(w) = i_0, \cdots, \\ &\quad \zeta_{n+m}(\omega) = i_m, \zeta_{n-1}, \cdots) \\ &= \begin{cases} p_{i_m j_m}, & i_1 = j_0, \cdots, i_{n+m} = j_{m-1}; \\ 0, & \text{其他}. \end{cases} \end{aligned}$$

即 $\{\eta_n(\omega)\}$ 仍是一个有限状态马氏链, 而且

$$\tilde{\Omega}_k \triangleq \{\omega: \eta_0(\omega) \in H_k^{m+1}\} = \{\omega: \eta_n(\omega) \in H_k^{m+1}\}$$

对 $\{\eta_n\}$ 是不变集, 而且 $\tilde{\Omega}_k$ 与 $\Omega_k$ 概率为 1 地相等, 从而 $\Omega_k$ 也是 $\{\eta_n\}$ 的不变集. 于是对 $\{\eta_n(\omega)\}$ 仍可再用上面所得结果. 注意到

$$\begin{aligned} P(w: \eta_0(\omega) = (i_0, \cdots, i_m)) &= p(w: \zeta_0(\omega) = i_0, \cdots, \zeta_m(\omega) = i_m) \\ &= \pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{m-1} i_m} \end{aligned}$$

与

$$\begin{aligned} P(\omega: \eta_0(\omega) = (i_0, \cdots, i_m) \mid \Omega_k) &= \pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{m-1} i_m} \Big/ \sum_{j\in H_k} \pi_j \\ &= \pi_{i_0}^{(k)} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{m-1} i_m}, \end{aligned}$$

将 (7.11) 与 (7.13) 用于 $\{\eta_n\}$ 及它所对应的 $g$ 就以概率为 1 地得到

$$\lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{n} g(\theta^l\omega) = \bar{g}(\omega)$$

$$= \sum_{k=1}^{s} I_{\Omega_k}(\omega) \sum_{j_0,\cdots,j_m} P(\eta_0(\omega) = (j_0,\cdots,j_m)\,|\,\Omega_k) f(j_0,\cdots,j_m)$$

$$= \sum_{k=1}^{s} I_{\Omega_k}(\omega) \sum_{j_0,\cdots,j_m} \pi_{j_0} p_{j_0 j_1} p_{j_1 j_2} \cdots p_{j_{m-1} j_m} f(j_0,\cdots,j_m)$$

与$\bar{g}(\omega)=\bar{g}(\theta\omega)$(也以概率为 1 地成立). 并且还有

$$Eg(\omega) = E\bar{g}(\omega).$$

进而还有：当 $\omega\in\Omega_k$ 时，

$$\sum_{j_0,\cdots,j_m} P(\eta_0(\omega) = (j_0,\cdots,j_m)\,|\,\Omega_k) f(j_0,\cdots,j_m)$$

$$= E(f(\eta_0(\omega))\,|\,\Omega_k) = E(g(\omega)\,|\,\Omega_k)).$$

**2. Birkhoff 遍历定理**

上段中我们得到的关于有限状态平稳马氏链的结果对于一般的保测映射及平稳序列仍然成立，这就是著名的 Birkhoff 遍历定理. 下面我们给出这个一般定理的阐述及应用，而略去其证明.

**定理 7.2(Birkhoff 遍历定理)** 设 $\theta$ 是概率空间$(\Omega,\mathscr{F},P)$上的一个保测映射，$g(\omega)$是$(\Omega,\mathscr{F},P)$上一个具有有限数学期望的随机变量，则

1) 极限

$$\lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{n} g(\theta^l\omega) = \bar{g}(\omega); \tag{7.15}$$

以概率为 1 地成立；

2) $\bar{g}(\theta\omega)=\bar{g}(\omega)$以概率为 1 地成立；

3) $Eg(\omega)=E\bar{g}(\omega)$.

其实定理 7.2 的结论正是(7.11)～(7.13)的翻版. 这里的 $g(\omega)$可以比 $f(\omega_0,\omega_1,\cdots,\omega_m)$这种形式更为一般.

**推论** 若 $\Omega_0\subset\Omega$ 满足 $P(\theta^{-1}\Omega_0\backslash\Omega_0)=P(\Omega_0\backslash\theta^{-1}\Omega_0)=0$，则对于 $\Omega_0$ 的示性函数 $I_{\Omega_0}$，有

$$E(g(\omega)I_{\Omega_0}) = E(\bar{g}(\omega)I_{\Omega_0}). \tag{7.18}$$

**注**：用测度论的语言来讲，(7.18)就是说$\bar{g}(\omega)$是$g(\omega)$对全体不变集组成的事件体($\sigma$-代数)的条件期望.

**推论证明** 令$\tilde{g}(\omega)=g(\omega)I_{\Omega_0}(\omega)$. 于是以概率为1地有

$$\frac{1}{n}\sum_{k=1}^{n}\tilde{g}(\theta^k\omega)=\frac{1}{n}\sum_{k=1}^{n}g(\theta^k\omega)I_{\Omega_0}(\theta^k\omega)=\frac{1}{n}\sum_{k=1}^{n}g(\theta^k\omega)I_{\Omega_0}(\omega)$$

(其中 $I_{\Omega_0}(\theta^k\omega)=I_{\theta^{-k}\Omega_0}(\omega)=I_{\Omega_0}(\omega)$ 以概率为1成立，因为 $\theta^{-k}\Omega_0$，$\theta^{-k+1}\Omega_0,\cdots,\theta^{-1}\Omega_0$ 与 $\Omega_0$ 都只差一个概率为0的集合(事件). 从而它们的示性函数当然也是以概率为1相等的). 于是由定理7.2，我们得到

$$\bar{\tilde{g}}(\omega)\triangleq\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\tilde{g}(\theta^k\omega)=(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}g(\theta^k\omega))I_{\Omega_0}(\omega)$$

$$=\bar{g}(\omega)I_{\Omega_0}(\omega).$$

再由 $E\bar{\tilde{g}}(\omega)=E\tilde{g}(\omega)$，便推出有

$$E(\bar{g}(\omega)I_{\Omega_0}(\omega))=E\tilde{g}(\omega)=E(g(\omega)I_{\Omega_0}(\omega)).$$

下面我们研究$\bar{g}(\omega)$何时与初状态无关，而且是非随机的. 先给出不变集与遍历平稳分布的一般概念：

**定义 7.3(不变集)** 对事件(集合)$A\in\mathscr{F}$，称 $A$ 对保测映射 $\theta$ 是不变的，如果 $P(A\backslash\theta^{-1}A)=P(\theta^{-1}A\backslash A)=0$.

**定义 7.4** 设有概率空间$(\Omega,\mathscr{F},P)$及其上的保测映射 $\theta$. 称概率 $P$ 对 $\theta$ 是遍历的，如果 $\theta$ 的不变集的概率非0则为1(也称 $\theta$ 为 $P$ 遍历的). 如令 $\xi_n(\omega)=\xi(\theta^n\omega)$，这时也称$\{\xi_n\}$是**遍历**的.

**定理 7.3** 在定理7.1中，要使对于任意有有限数学期望的

随机变量 $g$ 都以概率为 1 地有 $\bar{g}(\omega)=Eg(\omega)$ 当且仅当 $\theta$ 是 $P$ 遍历的. 这时

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}g(\theta^k\omega)=Eg(\omega). \tag{7.19}$$

这意味着沿轨道的时间平均等于空间的概率平均值.

**证明** (1) 当 $P$ 对 $\theta$ 遍历时,由于 $P(\omega:\bar{g}(\omega)=\bar{g}(\theta\omega))=1$,我们有

$$P(\omega:\bar{g}(\omega)\leqslant x<\bar{g}(\theta\omega))=P(\omega:\bar{g}(\theta\omega)\leqslant x<\bar{g}(\omega))=0.$$

可见 $\{\omega:\bar{g}(\omega)\leqslant x\}$ 是一个不变集. 由于 $P$ 对 $\theta$ 遍历,就有

$$P\{\omega:\bar{g}(\omega)\leqslant x\}=0\ 或\ 1.$$

令 $y_0=\inf(x:P(\omega:\bar{g}(\omega)\leqslant x)=1)$,则

$$P(\omega:\bar{g}(\omega)\leqslant y_0)=1,\quad P(\omega:\bar{g}(\omega)\leqslant y)=0\quad (y<y_0),$$

可见

$$P(\omega:\bar{g}(\omega)=y_0)=\lim_{\varepsilon\downarrow 0}P(\omega:y_0-\varepsilon<\bar{g}(\omega)\leqslant y_0)=1,$$

而且以概率为 1 地有 $y_0=E\bar{g}(\omega)=Eg(\omega)$.

(2) 设任意的具有有限数学期望的 $g$ 都使 $\bar{g}(\omega)=Eg(\omega)$,现在来证明 $P$ 对 $\theta$ 遍历. 设 $\Lambda$ 是 $\theta$ 的不变集,令它的示性函数为 $g(\omega)=I_\Lambda(\omega)$. 由于 $\theta$ 是保测映射,就有

$$\begin{aligned}P((\theta^{-n}\Lambda)\backslash(\theta^{-n+1}\Lambda))&=P((\theta^{-n+1}\Lambda)\backslash(\theta^{-n+2}\Lambda))=\cdots\\&=P((\theta^{-1}\Lambda)\backslash\Lambda)=0.\end{aligned}$$

$$P((\theta^{-n+1}\Lambda)\backslash(\theta^{-n}\Lambda))=\cdots=P(\Lambda\backslash(\theta^{-1}\Lambda))=0.$$

令 $\Omega_n=\{\omega:g(\theta^n\omega)=g(\theta^{n-1}\omega)\}$ $(n=1,2,\cdots)$,于是

$$\begin{aligned}P(\Omega_n)&=1-P\{\omega:g(\theta^n\omega)\neq g(\theta^{n-1}\omega)\}\\&\geqslant 1-[P((\theta^{-n+1}\Lambda)\backslash(\theta^{-n}\Lambda))+P((\theta^{-n}\Lambda)\backslash(\theta^{-n+1}\Lambda))]\\&=1.\end{aligned}$$

再令 $\Omega_0 = \{\omega: g(\theta^n \omega) = g(\theta^{n-1}\omega), n = 1, 2, \cdots\} = \bigcap_n \Omega_n$，就有

$$P(\Omega_0) \geqslant 1 - \sum_n (1 - P(\Omega_n)) = 1.$$

另一方面，在 $\Omega_0$ 上($P(\Omega_0) = 1$)有

$$\frac{1}{n} \sum_{k=1}^{n} g(\theta^k \omega) = g(\omega),$$

可见

$$g(\omega) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} g(\theta^k \omega) = g(\omega) = I_A(\omega)$$

以概率为 1 地成立. 由所设就得到

$$I_A(\omega) = g(\omega) = E\, g(\omega) = P(A),$$

即 $P(A) = 0$ 或 1. 由此可见 $P$ 对 $\theta$ 遍历.

从定理 7.3 反过来看有限状态的平稳马氏链，可以看出平稳马氏链的平稳初分布$\pi$集中在一个不可约常返状态类时，它对应于遍历的情形；反之当$\pi$在两个或两个以上不可约常返类上都有正测度时，例如设在 $H_1$ 与 $H_2$ 上$\pi$均有正测度，于是

$$\Omega_i = \{\omega: \zeta_n(\omega) \in H_i, 对\ \forall\ n \geqslant 0\} \quad (i = 1, 2)$$

均为不变集，而且

$$\begin{aligned} P(\Omega_i) &= P(\{\omega: \zeta_n(\omega) \in H_i, 对\ \forall\ n \geqslant 0\}) \\ &= P(\{\omega: \zeta_0(\omega) \in H_i\}) = \sum_{j \in H_i} \pi_j > 0 (i = 1, 2). \end{aligned}$$

于是

$$0 < P(\Omega_1) \leqslant 1 - P(\Omega_2) < 1.$$

可见这时 $P$ 对 $\theta$ 是不遍历的. 从而可见有限状态平稳马氏链遍历当仅当它的初分布只集中在一个常返类上.

### 3. Birkhoff 遍历定理的应用

通常许多随机系统如果它的运转机制是时间不变的，那么经过充分长的时间，与这个随机体系相关的一些随机过程就可近似

地认为是平稳过程.这一点可以从有限状态马氏链的特例来理解,当运转时间充分长后(即 $t\geqslant T\gg 1$),可认为 $\zeta_T(\omega)$ 遵从平稳分布(不变概率分布),$\{\zeta_{t+T}(\omega):t\geqslant 0\}$ 可近似地当作平稳过程.对于一个遍历的平稳序列 $\zeta=\{\zeta_n(\omega)\}$ 及 $g(\omega)=f(\zeta_0(\omega),\cdots,\zeta_m(\omega))$,由于概率为 1 地有

$$Eg(\omega)=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}g(\theta^k\omega) \tag{7.20}$$

$$=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}f(\zeta_k(\omega),\zeta_{k+1}(\omega),\cdots,\zeta_{k+m}(\omega)).$$

$\zeta$ 的一些重要的统计参数,如均值、方差、相关函数、协方差函数都可由 $\zeta$ 的一个现实(一条轨道)在足够长的时间中的取值来估计.例如其均值

$$\mu\triangleq E\zeta_0(\omega)=E\zeta_n(\omega)=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\zeta_k(\omega)$$

$$\approx\frac{1}{n}\sum_{k=1}^{n}\zeta_k(\omega)\quad(n\text{ 充分大}),$$

相关函数

$$R(\tau)=E(\zeta_{t+\tau}\zeta_t)=E(\zeta_\tau\zeta_0)=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\zeta_k(\omega)\zeta_{k+\tau}(\omega)$$

$$\approx\frac{1}{n}\sum_{k=1}^{n}\zeta_k(\omega)\zeta_{k+\tau}(\omega)\quad(n\gg 1).$$

对连续时间参数的平稳过程,由 Birkhoff 遍历定理也可得到相应的遍历定理.由于严格地叙述一般定理要用到许多测度论的工具,这里我们给出一个大大加强条件的定理,并直观地说明怎样由 Birkhoff 定理去推导这个定理.

**定理 7.4** 设 $\zeta=\{\zeta_t(\omega):t\geqslant 0\}$ 是一个平稳过程,$\zeta_t(\omega)$ 对 $t$,概率为 1 地连续.又设 $f(x)$ 是一个有界分段连续函数.则

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f(\zeta_t(\omega))\mathrm{d}t=\bar{g}(\omega)$$

概率为 1 地存在,而且 $E\bar{g}(\omega)=Ef(\zeta_0(\omega))$,同时 $\bar{g}(\theta^t\omega)=\bar{g}(\omega)$

概率为1的成立$(\forall\ x)$.

**推导说明** 令 $\eta_n(\omega)=\int_{n-1}^{n} f(\zeta_t(\omega))\mathrm{d}t$，于是$\{\eta_n(\omega)\}$是平稳序列，而且 $E|\eta_n(\omega)|<\infty$，从而由 Birkhoff 遍历定理可以得到

$$\lim_{N\to\infty}\frac{1}{N}\int_0^N f(\zeta_t(\omega))\mathrm{d}t=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\eta_k(\omega)\triangleq g(\omega)$$

概率为1地存在，并且

$$E\,g(\omega)=E\eta_1(\omega)=E\int_0^1 f(\zeta_s(\omega))\mathrm{d}s=Ef(\zeta_0(\omega)).$$

再注意到$|f(x)|\leqslant K<+\infty$（$f$ 有界），就可以得到：对 $N\leqslant T<N+1$，

$$\left|\frac{1}{N}\int_N^T f(\zeta_t(\omega))\mathrm{d}t\right|\leqslant\frac{1}{N}\int_N^{N+1}|f(\zeta_t(\omega))|\mathrm{d}t\leqslant\frac{K}{N}\to 0\quad(N\to\infty),$$

于是概率为1地有

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f(\zeta_t(\omega))\mathrm{d}t=g(\omega).$$

**例 7.1** 后面在第九章中定义 9.3 将给出的平稳 OU 过程 $\zeta=\{\zeta_t(\omega):t\geqslant 0\}$就满足本定理条件，并且以概率为1地成立

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f(\zeta_t(\omega))\mathrm{d}t=Ef(\zeta_0(\omega))=\int_{-\infty}^{+\infty}\frac{e^{-\frac{x^2}{2\gamma}}}{\sqrt{2\pi\gamma}}f(x)\mathrm{d}x.\ \blacksquare$$

事实上，定理 7.4 中关于 $f$ 的有界连续性条件及 $\zeta_t(\omega)$对 $t$ 的连续性条件均可大大减弱，而只要(7.21)左端的积分有定义，而且 $E|\zeta_t(\omega)|<\infty$即可. 于是对 OU 过程，由于它是遍历的，我们还有

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T\zeta_t(\omega)\mathrm{d}t=0,$$

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T\zeta_t(\omega)\zeta_{t+\tau}(\omega)\mathrm{d}t=E(\zeta_{t+\tau}\zeta_t)=R(\tau)$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{e^{-\frac{x^2}{2\gamma}}}{\sqrt{2\pi\gamma}}\cdot\frac{\exp\left\{-\frac{(y-xe^{-\beta\tau})^2}{2\gamma(1-e^{-2\beta\tau})}\right\}}{\sqrt{2\pi\gamma(1-e^{-2\beta\tau})}}xy\,\mathrm{d}x\mathrm{d}y$$

$$= \int_{-\infty}^{+\infty} \frac{e^{-\frac{x^2}{2\beta}}}{\sqrt{2\pi\beta}} x^2 e^{-\beta\tau} dx = \gamma e^{-\beta\tau}.$$

## §7.3 遍历论中的一些基本概念

本节介绍遍历论中的一些基本概念，它们不仅在遍历论中很重要，而且对数理统计、随机过程论都有重要意义. 对统计物理与非线性科学，它也都是重要的基本概念.

### 1. 混合与遍历

前面，我们给出了遍历的定义. 但是这个定义中的条件，即不存在概率非 0 或 1 的不变集，是十分难于检查的. 下面的命题在某些时候较为易于检查.

**命题 7.5** $\theta$ 是$(\Omega, \mathscr{F}, P)$的保测映射，则 $\theta$ 遍历当且尽当对 $\forall A, B \in \mathscr{F}$，使

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P(\theta^{-k} A \cap B) = P(A)P(B). \qquad (7.23)$$

**证明** 1) 当(7.23)成立. 设 $A$ 是一个 $\theta$ 的不变集，取 $A=B=A$，于是有

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P(A \cap \theta^{-k} A) = P(A)^2,$$

即 $P(A)=P(A)^2 \Rightarrow P(A)=0$ 或 1. 可见 $\theta$ 对 $P$ 遍历.

2) 当 $\theta$ 对 $P$ 遍历，于是有

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I_A(\theta^k \omega) I_B(\omega) = EI_A(\omega) I_B(\omega) = I_B(\omega) P(A)$$

(概率为 1 地成立). 由此可以得到(这一步不严格)：

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} E(I_A(\theta^k \omega) I_B(\omega)) = EI_A(\omega) EI_B(\omega),$$

即$\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{k=1}^{n} P((\theta^{-k} A) \cap B) = P(A)P(B)$.

另一个遍历论的重要概念是混合，它的定义如下：

**定义 7.5（混合）** 设 $\theta$ 是 $(\Omega, \mathscr{F}, P)$ 上的保测映射，设 $P$ 对 $\theta$（或 $\theta$ 关于 $P$）**是混合的**，如果对 $\forall\ A, B \in \mathscr{F}$，都有

$$P(\theta^{-n}A \cap B) \to P(A)P(B) \quad (n \to \infty). \tag{7.24}$$

混合的直观意义是：反复用 $\theta$ 对 $\Omega$ 映射足够多次，它就应在 $P$ 的意义下充分均匀；因为 $\theta^{-n}A = \{\omega: \theta^n\omega \in A\}$，表示经过 $n$ 次 $\theta$ 映射进入 $A$ 的点集；若一开始在 $B$ 中放入标记物（如墨水），则从 $B$ 中出发，经 $n$ 次 $\theta$ 映射进入 $A$ 的点集就是 $\theta^{-n}A \cap B$，(7.24) 告诉我们当 $n$ 充分大，不管 $A$ 在那儿，这个集合的大小 $P(\theta^{-n}A \cap B)$ 近于和 $A$ 与 $B$ 的大小 $P(A)$、$P(B)$ 都成比例. 这不是意味着标记物到处近于均匀了吗？

(7.23) 也道明了遍历的直观含义：注意到

$$\frac{1}{n}\sum_{k=1}^{n} P(\theta^{-n}A \cap B) = E\left(\frac{\sum_{k=1}^{n} I_{\theta^{-k}A}(\omega) I_B(\omega)}{n}\right) = E(\nu_{AB}^n(\omega)),$$

其中 $\nu_{AB}^n(\omega) = N_{AB}^n(\omega)/n$，而 $N_{AB}^n(\omega) \triangleq \#\{k: \theta^k\omega \in A, \omega \in B\}$. (7.23) 即 $E\nu_{AB}^n(\omega) \approx P(A)P(B)$. $N_{AB}^n(\omega)$ 是从 $B$ 出发在前 $n$ 次 $\theta$ 的映射下，由 $B$ 出发访问 $A$ 的次数，$\nu_{AB}^n(\omega)$ 就是访问 $A$ 的频率. (7.24) 是说这个频率的期望（概率平均值）是与 $A$，$B$ 的大小成比例的. 也就是说，时间充分长后，按时间平均不管从那儿出发 $(\forall\ B)$ 它将在按 $P$ 平均的意义下均匀地访问 $\Omega$ 的各部分. 下面的命题告诉我们：混合是比遍历更强的性质.

**命题 7.6** 若 $\theta$ 对 $P$ 混合则 $\theta$ 对 $P$ 遍历.

**证明** 由于极限存在能推出平均极限存在而且与之相同，由 (7.23) 与 (7.24)，命题 7.6 显然成立.

**例 7.2** （平稳马氏链的遍历性与混合性）

设 $\xi$ 是有限状态平稳马氏链. 命题 7.5 告诉我们 $\xi$ 遍历的充要条件是它的平稳初分布恰在一个常返类上有负荷. 而再用命题 7.6，还可得到 $\xi$ 混合当且仅当它是非周期而且遍历的，我们略去

它的证明.

**说明**：命题 7.6 告诉我们，若 $\zeta$ 混合，则它必须遍历. 它只有一个常返类概率负荷不为 0，又由 § 3.3 中 4 段就知道这时 $\zeta$ 非周期是 $\lim\limits_{n\to\infty} p_{ij}(n)=\pi_j$ 的充要条件. 然而

$$\pi_i p_{ij}(n) = P(\zeta_0(\theta^n\omega) = j, \zeta_0(\omega) = i)$$

$$= P(\theta^{-n}A \cap B) \to P(A)P(B), \qquad (7.25)$$

其中 $A=\{\omega:\zeta_0(\omega)=j\}$，$B=\{\omega:\zeta_0(\omega)=i\}$. 于是若 $\zeta$ 有周期时，$p_{ij}(n)$ 没有极限，那么显然 $\theta$ 不混合，反之，如马氏链常返互通非周期，则由 (7.23) 可以得到对一般的 $A,B\in\mathscr{F}$，都有 (7.24) 成立(这里的证明要用到更多分析工具，这里我们略去它).

## 2. 熵与相对熵

熵的概念最早由 C. Shannon 在信息论中提出，用以刻画一个离散值随机变量或一个离散分布的不确定性的大小，或者说刻画知道此随机变量的取值所获得的信息量的大小. 先让我们从最简单的随机变量，一个随机事件 $A$ 的示性函数 $I_A(\omega)$(它只有 0 与 1 两个取值)出发来分析. 知道 $I_A(\omega)=1$(即 $A$ 发生)所获取的信息应该是 $A$ 发生的概率 $P(A)$ 的函数. 将此信息记为 $H(A)$，即有

$$H(A) = \varphi(P(A)). \qquad (7.26)$$

如果有人告诉你："美国梦幻队在奥运会胜了日本队"，你会觉得他没有给你多少信息，甚至认为他说了一句废话. 反之，如果有人告诉你一个从未输过的名将在一次比赛中名落孙山，这就使你觉得这一消息的信息量很大了. 原因是前一事件几乎概率为 1 地发生，而后一事件发生的概率极小. 由此，读者可以体会到：前面认为 $H(A)$ 是 $P(A)$ 的函数是十分合理的. Shannon 还对信息量作了如下规定：

1) $H(A)\geqslant 0$，$H(\Omega)=0$；

2) 当 $A$ 与 $B$ 独立，$H(A\cap B)=H(A)+H(B)$； (7.27)

3) 当 $P(A_n)\to P(B)$，$H(A_n)\to H(A)$.

这三项规定都与直观很一致：知道一个必然发生的事发生了，当然没有获得信息；同时知道两件不相干的事发生了，所获得的信息应是分别获得的信息之和；第三项假定有一点人为的技术因素，但也还是很自然的. 用(7.26)中的 $\varphi$ 来改写(7.27)的 1)～3)即有：

1° $\varphi$ 是 $[0,1]$ 区间(概率 $0\leqslant P(A)\leqslant 1$)上的一个连续函数，$\varphi(1)=0$.

2° $\varphi(pq)=\varphi(p)+\varphi(q)$.

3° $\varphi$ 连续.

由此可见 $\varphi(p^n)=n\varphi(p)$，也即 $\varphi(p)=\frac{1}{n}\varphi(p^n)$，于是 $\varphi(p^{\frac{m}{n}})=\frac{m}{n}\varphi(p)$，用 $\varphi$ 的连续性就得到对于 $0<\alpha\leqslant 1$，$\varphi(p^\alpha)=\alpha\varphi(p)$. 令 $\varphi(e^{-1})=c$，那么

$$\varphi(p)=\varphi(e^{-\log\frac{1}{p}})=\left(\log\frac{1}{p}\right)\varphi(e^{-1})\triangleq -c\log p,$$

其中 $c$ 显然不可能是 0，否则 $\varphi(p)=0$ 就无意义了. 这里的 $c$ 可看成是选取的单位常数.

综上所述，已知随机变量 $\xi(\omega)=I_A(\omega)$ 的结果所获得的信息量是 $H(A)=-c\log P(A)=-\log_a P(A)\left(取\ a\ 使\ c=\frac{1}{\log a}\right)$. 下面我们再从随机小数 $\xi(\omega)$($\in[0,1]$，它服从 $[0,1]$ 上的均匀分布)来认识 Shannon 信息量的定义，将 $\xi(\omega)$ 写成二进制小数形式：

$$\xi(\omega)=\sum_{n=1}^{\infty}a_n(\omega)2^{-n}\quad(a_n(\omega)=0,1).$$

于是 $\{a_n(\omega)\}$ 是相互独立同分布的随机变量列，且 $P(\omega:a_n(\omega)=0)=P(\omega:a_n(\omega)=1)=\frac{1}{2}$. 因此，知道了一位二进制小数的取值所获得的信息量是 $-\log_a\left(\frac{1}{2}\right)=\log_a 2$. 若取 $a=2$，即 $c=\frac{1}{\log_a 2}$，也就有 $\log_a 2=1$，也就是以由一位二进制小数取值所获得的信息量取为 1(个单位)，那么知道两位二进制小数所获信息量是 $-\log_2\frac{1}{4}=2$，

因为

$$P(\omega: a_n(\omega) = 0, a_{n+1}(\omega) = 1) = P(\omega: a_n(\omega) = 1, a_{n+1}(\omega) = 1)$$
$$= P(\omega: a_n(\omega) = 0, a_{n+1}(\omega) = 0)$$
$$= P(\omega: a_n(\omega) = 1, a_{n+1}(\omega) = 0) = \frac{1}{4},$$

进而可得知 $m$ 位二进制小数所获信息量应为：$-\log_2\left(\frac{1}{2^m}\right)=m$. 当然我们不能只顾及随机小数，所以在一般情形为了省写常数，我们取 $a=e, c=\frac{1}{\log e}=1$.

设一个随机试验 $\xi$ 可有 $M$ 个不同的结果：$A_1,\cdots,A_M$，它们发生的概率分别为：

$$P(A_i) = p_i, \quad \sum_i p_i = 1, p_i > 0 \quad (i = 1,\cdots,M).$$

于是知道试验结果是 $A_i$ 所获的信息量是 $-\log p_i$. 知道不同的试验结果所获信息量是不同的. 总的描述这一试验结果所获信息量 $H(\xi)$ 是知道不同试验结果所获信息量按概率的平均值（数学期望）：

$$H(\xi) = \sum_{i=1}^{M} p_i(-\log p_i) = -\sum_{i=1}^{M} p_i \log p_i.$$

Shannon 把 $H(\xi)$ 定义为此随机试验的熵. 这个随机试验也可理解为 $\Omega$ 的一个分割：

$$\Omega = \bigcup_{i=1}^{M} A_i, A_i \cap A_j = \psi \quad (i \neq j).$$

**定义 7.6（分割的熵**，随机变量（取可数个值）之熵）　设 $\Omega$ 有一个可数分割 $\xi:\{A_1,\cdots,A_M\}$，$P(A_i)=p_i(i=1,\cdots,M)$，则此分割 $\xi$ 的熵定义为

$$H(\xi) \triangleq -\sum_{i=1}^{M} p_i \log p_i \quad (M \leqslant +\infty) \tag{7.27}$$

（$H(\xi)$ 可以取 $+\infty$）. 具有可数个取值的随机变量 $\xi(\omega)\in\{a_1,\cdots,a_M\}$ 的熵定义为分割 $\{A_1,\cdots,A_M\}$ 之熵，其中

$$A_i = \{\omega:\xi(\omega) = a_i\},\quad i = 1,2,\cdots,M.$$

**定义 7.7(相对熵)** 设有分布 $P=\{p_i:i=1,2,\cdots\}$与 $Q=\{q_i:i=1,2,\cdots\}$,则分布 $P$ 相对于 $Q$ 的相对熵定义为

$$h(P,Q)\triangleq\begin{cases}\sum_i p_i\log\frac{p_i}{q_i}, & q_i=0\text{ 时必有 }p_i=0(\forall\ i);\\ +\infty, & \text{其他情形}.\end{cases}\tag{7.28}$$

**命题 7.7** $h(P,Q)\geqslant 0$,而且 $h(P,Q)=0$ 当且仅当 $p_i\equiv q_i(i=1,\cdots)$.

**证明** 不妨设 $h(P,Q)<+\infty$. 由于

$$g(t)\triangleq t-1-\log t>0\quad(t>0,\neq 1).$$

因为

$$g'(t)=\begin{cases}1-\frac{1}{t}>0, & \text{当 }t>1;\\ 1-\frac{1}{t}<0, & \text{当 }t<1,\end{cases}$$

$$g(1)=0.$$

可见 $g(t)>g(1)=0\quad(t\neq 1)$. 取 $t=\frac{q_i}{p_i}$,于是 $\log\frac{q_i}{p_i}\leqslant\frac{q_i}{p_i}-1$ (而且等号仅当 $p_i=q_i$ 时成立). 从而

$$\sum_i p_i\log\frac{p_i}{q_i}\geqslant\sum_i p_i\left(1-\frac{q_i}{p_i}\right)=\sum_i(p_i-q_i)=0.$$

而且上式中等号仅当 $p_i=q_i(i=1,2,\cdots)$成立.

**命题 7.8.1** 设 $\xi(\omega)$是具有 $M$ 个不同取值的随机变量,$\xi$ 取它们的概率分别为 $p_1,\cdots,p_M$. 则

$$H(\xi)=-\sum_{i=1}^{M}p_i\log p_i\leqslant\log M\left(=-\sum_{i=1}^{M}\frac{1}{M}\log\frac{1}{M}\right).$$

即当 $\xi$ 以相同概率$\left(\frac{1}{M}\right)$取不同的 $M$ 个值时,$\xi$ 的熵最大.

**证明**

$$-\sum_{i=1}^{M} p_i \log p_i - \log M = -\left(\sum_{i=1}^{M} p_i (\log p_i - \log \frac{1}{M})\right)$$

$$= -\sum_{i=1}^{M} p_i \log \frac{p_i}{\frac{1}{M}} \qquad (\text{取 } q_i = \frac{1}{M})$$

$$= -\sum_{i=1}^{M} p_i \log \frac{p_i}{q_i} \leqslant 0 \qquad (\text{由命题 7.7}).$$

命题 7.7 还有一个重要意义，就是它说明相对熵表述出 $P$ 与 $Q$ 的差别：当 $P=Q$ 时，$h(P,Q)=0$. 而 $p_i$ 与 $q_i$ 均十分接近时，$h(P,Q)$就很小. 从而 $h(P,Q)$可以看成 $P$ 与 $Q$ 之间的准距离. 我们之所以说是准距离，是因为它不对称，也不满足三角形不等式. 所以不能满足距离公理.

对于有概率分布密度的随机变量 $\xi(\omega)$，它取某一个定值的概率总是零，所以直接按(7.28)定义就无意义. 但是，我们可以按(7.28)的思想做一点变化，定义：

**定义 7.8** 设 $\xi(x)$是以 $p(x)$为概率分布密度的随机变量，其熵定义为(或称分布密度 $p(x)$的熵)：

$$H(\xi) \triangleq -\int p(x)\log p(x)\mathrm{d}x. \qquad (7.27)'$$

分布密度 $p(x)$对 $q(x)$的相对熵定义为 $\int p(x)\log \frac{p(x)}{q(x)}\mathrm{d}x$.

命题 7.8.1 告诉我们有限的离散分布中以均匀情形的熵最大. 那么有限区间上具有密度的分布何者熵最大呢?非有限区间的情况如何呢? Boltzmann 多年前曾给出解决这个问题的思想.

**命题 7.8.2** 在具有有限的数学期望 $E\xi=\mu$ 的条件下取值于$\{x_1,x_2,\cdots,x_k,\cdots\}$的各种离散分布中，以分布为

$$p_k = P(\xi(\omega)=x_k) = C\mathrm{e}^{-\frac{x_k}{\lambda}}\left(C=\left(\sum_k \mathrm{e}^{-\frac{x_k}{\lambda}}\right)^{-1}\right)$$

的熵最大($\lambda$ 依赖于 $\mu$)；其熵为$\frac{\mu}{\lambda}-\log C$.

**证明** $H(\xi_\lambda) = -\sum_k Ce^{-\frac{x_k}{\lambda}}\left[\log C - \frac{x_k}{\lambda}\right] = \frac{\mu}{\lambda} - \log C$. 而且对任一分布$\{p_k\}$有

$$0 \leqslant \sum_k p_k \log \frac{p_k}{Ce^{-\frac{x_k}{\lambda}}} = \sum_k p_k \log p_k + \frac{1}{\lambda}\sum_k p_k x_k - \log C$$

$$= \frac{\mu}{\lambda} - \log C + \sum_k p_k \log p_k.$$

从而

$$H(\xi) = -\sum_k p_k \log p_k \leqslant \frac{\mu}{\lambda} - \log C.$$

**命题 7.8.3** 在均值为$\mu$、方差为$\sigma^2$的取值于$\{x_k: k=1,2,\cdots\}$的全部离散分布中以分布

$$p_k = Ce^{-\frac{(x_k-\alpha)^2}{\beta^2}} \quad \left(C = \left[\sum_k e^{-\frac{(x_k-\alpha)^2}{\beta^2}}\right]^{-1}\right)$$

的熵最大($(\alpha,\beta^2)$依赖于$(\mu,\sigma^2)$). 其熵为$\frac{\sigma^2+(\mu-\alpha)^2}{\beta^2} - \log C$.

**证明** 作为习题,读者自证.

**命题 7.8.4** 指数分布是在$[0,+\infty)$上具有同均值的有密度的分布中熵最大的;$N(a,\sigma^2)$是在$R^1$,具有均值$a$与方差$\sigma^2$的具有密度的分布中熵最大的;而均匀分布是取值于有限区间$[a,b]$上的有密度中,熵达最大的. 其最大熵分别为

$$1 - \log\lambda, \quad \frac{1}{2}(1 + \log 2\pi) + \log\sigma, \quad \log(b-a).$$

**证明** 这里关键的问题是先建立不等式

$$\int \rho(x)\log\frac{\rho(x)}{p(x)}\mathrm{d}x \geqslant 0$$

及利用$\int\rho(x)\mathrm{d}x = \int p(x)\mathrm{d}x = 1$. 其他证明类似于命题 7.8.1~7.8.3.

下面我们用熵的概念来解决一个著名的游艺难题:称珠问题.

**例 7.3** 有珠 13 颗,外形完全相同,其中有一颗是假珠,比其

他珠轻，现有天平一架，但无法码，试问至少要用多少次才能找出假珠，应怎样去称？

**分析** 由于天平无法码，所以所谓称，无非是将要称的珠分为三组，其中两组有相同个数（设为 $k$ 个），分别放上天平两端，看是否天平是平的. 若是平的，说明假珠在剩下的那组中，否则假珠在天平上. 称一次就对应一个随机试验，这个试验就是如下的一个分割：总的个数为 $m$ 个珠，分为三组：$k$ 个，$k$ 个，$m-2k$ 个. 由于 $m$ 个中任意一个是假珠的概率都是 $\frac{1}{m}$，于是称一次这个随机试验共有三个结果：假珠在天平左边，右边和剩下那组中. 我们把这三种情形分别记成 $A_1, A_2, A_3$. 这时

$$
\begin{aligned}
H(\xi) &= -(p(A_1)\log p(A_1) + p(A_2)\log p(A_2) + p(A_3)\log p(A_3)) \\
&= -\left(\frac{2k}{m}\log\frac{k}{m} + \left(1-\frac{2k}{m}\right)\log\left(1-\frac{2k}{m}\right)\right) \\
&\leqslant \log 3.
\end{aligned}
$$

如果总共称 $N$ 次，记第 $l$ 次称所有可得结果为 $A_1^l, A_2^l, A_3^l$，于是总的结果共有 $3^N$ 个. 我们重新把 $N$ 次称珠作为一个新的随机试验，将这一随机试验的熵记为 $H_N$，那么它等于

$$
H_N = -\sum_{i=1}^{3^N} p_i \log p_i \leqslant \log 3^N = N\log 3,
$$

其中 $p_i$ 是第 $i$ 种可能的结果出现的概率. 所以称 $N$ 次所能获得的信息不能超过 $N\log 3$. 然而要从 13 个球中辨出某一个轻珠所获信息应为 $-\log\frac{1}{13}=\log 13\left(\text{因为某一个特定珠是假的概率是}\frac{1}{13}\right)$. 由此可见要在 $N$ 次称珠中找出假珠，所获信息量（熵）应大于或等于 $\log 13$. 即要求：$N\log 3 \geqslant H_N \geqslant \log 13$. 即 $N \geqslant \frac{\log 13}{\log 3} = \log_3 13 > \log_3 9 = 2$. 这就给出了应称次数的一个下界，即至少要称 3 次.

下面我们来设计称珠方案. 设第 $l$ 次称时是从 $m_l$ 个珠中，选出两组 $k_l$ 个珠放上天平左边与右边，剩下 $m_l-2k_l$ 个珠. 显然 $m_l$

$=13$，而 $m_2,m_3$ 可根据第一次称的结果而定. 第一次称的最好方案是获取最大信息的方案，即使

$$H_1=-\left[\frac{2k_1}{13}\log\frac{k_1}{13}+\left(1-\frac{2k_1}{13}\right)\log\left(1-\frac{2k_1}{13}\right)\right]$$

尽量接近 log3 的方案，即要 $\frac{k_1}{13}$ 尽量接近 $\frac{1}{3}$（参见命题 7.7 的证明）. 所以 $k_1$ 应取为 4. 这样，第一次称应为取出两组，每组各 4 个珠上天平去称. 可能得到的结果是：

1）天平两端平，假珠在剩下 5 个中. 于是这时第二次只要对 $m_2=5$ 去重复上面的分析.

2）天平两端不平，则假珠应在天平上的轻端的 4 个珠中. 这时 $m_2=4$.

对 1)的情形第二次称是将 5 个珠分为三组，上天平的个数 $k_2\leqslant 2$（因为要使 $2k_2\leqslant 5$）. 于是此试验的熵是

$$\left(\frac{2k_2}{m_2}\log\frac{k_2}{m_2}+\frac{m_2-2k_2}{m_2}\right)\log\left(1-\frac{2k_2}{m_2}\right).$$

好的称法是使它接近于 log3，即 $\frac{k_2}{m_2}=\frac{k_2}{5}\sim\frac{1}{3}$. 于是应取 $k_2=2$，即将第一次称剩下的 5 个珠中选两组各 2 个放上天平. 如果这时天平两端平，则假珠是剩下的那个；如果天平两端不平，则假珠是轻端的两个之一，而其他三个均为好珠. 只要将轻端的两个珠任选一个与好珠上天平去比，如得到天平两端平，则轻端两个珠中没有被选上天平的那个就是假珠，否则在天平上的那个是假的.

如果第一次称天平两端不平，则轻端那 4 个珠中之一是假的，它比从 5 个珠中找出假的更容易. 可如前面的方法称得假珠. 综上分析可知称三次就足够了.

对于未知假珠的轻重要在 13 个珠中称出 1 个假珠来，按上面类似地估计的所称次数有下限 $k\geqslant 3$（即 $k\log3\geqslant\log26$，因为未知假珠轻重时我们是要从 26 种不同结果：第 $j$ 个珠轻于正常、第 $j$ 个珠重于正常（$j=1,2,\cdots,13$）中找出所发生的答案）. 但是这只是必

要条件．实际上 $k=3$ 是不够的，我们再仔细分析一下．事实上第一次称完后剩下需获信息是：当 $A_1^i(i=1,2)$ 发生，则还需确定 $13-2k_1$ 个珠中的那个假珠，于是 $(k-1)\log 3 \geqslant \log 2(13-2k_1)$；当 $A_1^3$ 发生，则还需要确定 $2k_1$ 个珠中所含的那个假珠，因而 $(k-1)\log 3 \geqslant \log 2k_1$（因为这时未上天平的珠一定全是真珠，我们已有真珠样品了）．

综合两种情形的分析可以得到

$$(k-1)\log 3 \geqslant (\log 2(13-2k_1)) \vee (\log 2k_1).$$

由于当 $k_1 \geqslant 5$ 时，$\log 2k_1 \geqslant \log 10$；$k_1 \leqslant 4$ 时，$\log 2(13-2k_1) \geqslant \log 10$，从而必须有 $k \geqslant 1+\log_3 10$（即 $k \geqslant 4$），故至少要称 4 次才行．事实上称 4 次就够了，但我们略去具体的称法．

**例 7.4** 神经网络的学习律的功能分析．

设有一个 $N$ 个神经元，每个可取二值：$-1$ 或 1，神经元之间可以有对相互作用（见 § 4.4.1 段）．要使这神经网络学会信息 $\underline{y}=(y_1,\cdots,y_N)$，也就是要使 $\underline{y}$ 变成此网络的一个吸引子，常见的学习方法是多次逐步改变代表相互作用的联结系数 $\{W_{ij}\}$，使得最终 $\underline{y}$ 能变成一个吸引子．在每一次改变中有几个最有名的学习律，即选取 $W_{ij}$ 的改变量 $\Delta W_{ij}$ 的方式如下：

Hebb 律 $\quad \Delta W_{ij} = \eta y_i y_j$，

Hopfield 学习律 $\quad \Delta W_{ij} = \eta(2y_i-1)(2y_j-1)$.

从相对熵的观点，可以分析出这两个著名学习律的优点．其实，学习的目的是希望改变以后的联结系数 $W_{ij}'=W_{ij}+\Delta W_{ij}$（把原来的联结系数 $W_{ij}$ 改为新的 $W_{ij}'$），能使从 $\underline{y}$ 出发的神经网络所得到的输出尽量与 $\underline{y}$ 接近（这就最接近于使 $\underline{y}$ 是吸引子）．然而从 $\underline{y}$ 出发的实际输出按（4.23）应是以 $P_y$ 为分布的随机变量 $\xi(\omega)$，其中

$$P_y(\underline{z}) = P(\xi(\omega)=\underline{z}) = P_{\underline{y}\underline{z}}$$

$$= \prod_{i=1}^{N} \{1+\exp[-\beta(\sum_{j=1}^{N} W_{ij}'y_j - \theta_i)z_i]\}^{-1}$$

(见(4.23)式)，要使 $\xi(\omega)$ 与 $y$ 近，就是要 $P_y$ 与集中在 $y$ 的一点分布 $\delta_y$ 接近，也就是相对熵

$$h(\delta_y, P_y) = \sum_z \delta_y(z)\log\frac{\delta_y(z)}{P_{yz}}$$

$$= -\sum_{i=1}^{N}\log\{1+\exp(-\beta(\sum_{j=1}^{N}(W_{ij}+\Delta W_{ij})y_j - \theta_i)z_i)\}$$

(7.30)

尽量接近于0，其中 $\delta_y = \{\delta_y(z)\}$，而

$$\delta_y(z) = \begin{cases} 1, & y = z; \\ 0, & y \neq z. \end{cases}$$

由于(7.30)中和号各项皆为正的，我们就知道和号下各项皆达其最大必然使(7.30)达到最小.

在各种学习律中对修改量总有一定的界限，不可以取遍 $(-\infty, +\infty)$. 事实上(假定 $\theta_i = 0$)，

1) 当限制 $0 \leqslant \Delta W_{ij} \leqslant \eta$ 时，$\Delta W_{ij} = \eta y_i y_j$，

2) 当限制 $|\Delta W_{ij}| \leqslant \eta$ 时，$\Delta W_{ij} = \eta(2y_i - 1)(2y_j - 1)$

均为使 $h(\delta_y, P_y)$ 达到最小的修改量 $\{\Delta W_{ij}\}$.

由此可见，Hebb 律这一在生物学中早就发现的自然界常存的学习律原来是经过长期自然选择而在条件 $0 \leqslant \Delta W_{ij} \leqslant \eta$ 下达到最优的学习律. 而 Hopfield 的学习律则是在对称的要求 $|\Delta W_{ij}| \leqslant \eta$ 下，达到最优的学习律. 在习题中我们还可以看到其他几种常见的学习律也是在某种意义下最优的学习律.

在实际的"学习"操作中，首先要选一个初始联结 $\{W_{ij}^{(0)}\}$，然后按我们事先要求的学习律(例如 Hebb 律或 Hopfield 律)反复进行递推：

$$W_{ij}^{(n+1)} = W_{ij}^{(n)} + \Delta W_{ij}^{(n)},$$

到足够长的 $n$ 为止($W_{ij}^{(0)}$ 的选取可以有一定的自由度).

### 3. 条件熵与保测映射的熵

若对概率空间有两个有限分割 $\pi_1$ 与 $\pi_2$,

$$\pi_1:\{A_1,A_2\cdots,A_N\}.\Omega=\bigcup_{k=1}^{N}A_k,A_kA_j=\varnothing\quad(k\neq j),$$

$$\pi_2:\{B_1,B_2\cdots,B_M\},\Omega=\bigcup_{k=1}^{M}B_k,B_kB_j=\varnothing\quad(k\neq j).$$

**定义 7.8(条件熵)** 分割 $\pi_1$ 对 $\pi_2$ 的条件熵定义为

$$H(\pi_1/\pi_2)=H(\pi_1\vee\pi_2)-H(\pi_2).$$

其中 $\pi_1\vee\pi_2$ 是分割 $\{A_kB_l:k=1,\cdots,N,l=1,\cdots,M\}$.

事实上

$$\begin{aligned}H(\pi_1\vee\pi_2)&=-\sum_{k,l}P(A_kB_l)\log P(A_kB_l)\\&=-\sum_{k,l}P(A_kB_l)(\log P(B_l)P(A_k|B_l)).\end{aligned}$$

于是

$$\begin{aligned}&H(\pi_1/\pi_2)\\&\quad=-\sum_{k,l}P(B_l)P(A_k|B_l)(\log P(B_l)P(A_k|B_l))\\&\qquad+\sum_{l}P(B_l)\log P(B_l).\\&\quad=-\sum_{k,l}P(B_l)P(A_k|B_l)\log P(A_k|B_l)\\&\quad=\sum_{l}P(B_l)\Big(-\sum_{k}P(A_k|B_l)\log P(A_k|B_l)\Big)\geqslant 0.\quad(7.31)\end{aligned}$$

$\pi_1\vee\pi_2$ 的含义是先做 $\pi_1$,再按 $\pi_2$ 将 $\pi_1$ 的每个子集再细分为:$\{A_kB_l:\ l=1,\cdots,M\}$.我们将此简称为连做 $\pi_1,\pi_2$ 分割.$\pi_2$ 对 $\pi_1$ 的条件熵是先做分割 $\pi_1$,再按 $\pi_2$ 细分所能增加的平均信息量.它也恰是在 $\pi_1$ 的每个事件 $A_k$ 的条件下,对更细的划分:$A_k=\bigcup_{l}A_kB_l$ 的分割的条件概率的熵,再按 $P(A_k)$ 作为权值的(概率)平均值.

**定义 7.9(随机序列的熵)** 设 $\xi_n(\omega)\in S=\{s_1,\cdots,s_N\}$,$(n=1,2,\cdots)$,则 $\{\xi_n(\omega)\}$ 的熵定义为

$$H(\xi)=\lim_{n\to\infty}\frac{1}{n}H_n(\xi_1,\cdots,\xi_n),$$

其中$H_n(\xi_1,\cdots,\xi_n)$是对分割$\{\{\omega:\xi_1(\omega)=s_{i_1},\cdots,\xi_n(\omega)=s_{i_n}\}\}$的熵.

设$\xi$是以$P_{(S\times S)}$为转移概率矩阵并以$\xi_0$为初分布的马氏链，则

$$H(\xi)=\lim_{n\to\infty}\frac{1}{n}H_n(\xi_0,\cdots,\xi_n)\quad(\text{由条件熵定义})$$

$$=\lim_{n\to\infty}\frac{1}{n}(H(\xi_0)+H(\xi_1|(\xi_0))+H(\xi_2|(\xi_0,\xi_1))+\cdots$$

$$+H(\xi_n|(\xi_0,\cdots,\xi_{n-1}))).$$

其中$\xi_k$与$(\xi_0,\cdots,\xi_{k-1})$分别理解为它们各自取不同的值的事件组成的分割. 于是

$$H(\xi_k|(\xi_0,\cdots,\xi_{k-1}))$$

$$=\sum_{s_{i_0},\cdots,s_{i_{k-1}}}P(\omega:\xi_0(\omega)=s_{i_0},\cdots,\xi_{k-1}(\omega)=s_{i_{k-1}})$$

$$\times H(\xi_k|\xi_0(\omega)=s_{i_0},\cdots,\xi_{k-1}(\omega)=s_{i_{k-1}}),$$

其中$H(\xi_k|\xi_0(\omega)=s_{i_0},\cdots,\xi_{k-1}(\omega)=s_{i_{k-1}})$表示$\xi_k$对于事件

$$\{\omega:\xi_0(\omega)=s_{i_0},\cdots,\xi_{k-1}(\omega)=s_{i_{k-1}}\}$$

的条件概率分布的熵. 又由于$\xi$是马氏链，上述条件概率即$\xi_k$对事件$\{\omega:\xi_{k-1}(\omega)=s_{i_{k-1}}\}$的条件概率，故而

$$H(\xi_k|\xi_0(\omega)=s_{i_0},\cdots,\xi_{k-1}(\omega)=s_{i_{k-1}})=H(\xi_k|\xi_{k-1}(\omega)=s_{i_{k-1}}).$$

由此

$$H(\xi_k|(\xi_0,\cdots,\xi_{k-1}))$$

$$=\sum_{s_{i_0},\cdots,s_{i_{k-1}}}P(\omega:\xi_0(\omega)=s_{i_0},\cdots,\xi_{k-1}(\omega)=s_{i_{k-1}})$$

$$\times H(\xi_k|\xi_{k-1}=s_{i_{k-1}})$$

$$=H(\xi_k|\xi_{k-1})=\sum_{i,j}\pi_i p_{ij}\log p_{ij}.$$

这样就有

$$H(\xi)=\lim_{n\to\infty}\frac{1}{n}(H(\xi_0)+H(\xi_1|\xi_0)$$
$$+H(\xi_2|\xi_1)+\cdots+H(\xi_n|\xi_{n-1}))$$
$$=-\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{N}\pi_i\log\pi_i+\lim_{n\to\infty}\frac{1}{n}\times n\{-\sum_{i,j}\pi_i p_{ij}\log p_{ij}\}$$
$$=-\sum_{i,j}\pi_i p_{ij}\log p_{ij}.$$

将这个概念推广到保测映射，我们定义：

**定义 7.10（保测映射的熵）** 设$(\Omega,\mathscr{A},P)$上有保测映射$\theta$，又设$\Pi:\Omega=\bigcup_{n=1}^{N}A_n$是$\Omega$的一个分割. 定义$\theta$对的熵为

$$H(\theta,\Pi)\triangleq\lim_{n\to\infty}\frac{1}{n}H(\Pi\vee(\theta^{-1}\Pi)\vee\cdots\vee(\theta^{-n}\Pi)),$$

其中$\theta^{-k}\Pi=\{\theta^{-k}A_n:n=1,2,\cdots,N\}$，而$\theta$的熵定义为

$$H(\theta)=\sup_{\Pi}H(\theta,\Pi).$$

上式右端上确界的范围取遍一切有限分割.

## 习　　题

1. 设$\xi=\{\xi_n:n=0,1,2,\cdots\}$是平稳序列，试证明$\{\eta_n\}$仍为平稳序列，其中

$$\eta_n=a_0\xi_n+a_1\xi_{n-1}\quad(n=1,2,\cdots).$$

2. 设$\xi(\omega)$是取值于$[0,2\pi]$上的均匀分布随机变量，则$\xi\triangleq\{\xi_t(\omega)\}$是平稳过程，其中

$$\xi_t(\omega)=\sin(\xi(\omega)+\theta t)\quad(t\geqslant 0).$$

3. 试证明：

1) 若$A$是一个不变集，则其余集$A^c$亦然

提示：注意$(\theta^{-1}A)^c=\theta^{-1}A^c$.

2) 若$A,B$皆为不变集，则$A\cap B,A\cup B$皆为不变集.

3) 若$A_n\uparrow A,A_n$均为不变集，则$A$亦然.

3. 在$(\Omega, \mathscr{F}, P)$上的保测映射$\theta$的全体不变集组成一个事件体($\sigma$代数)(记为$\mathscr{I}_\theta$).

4. 设$\xi(\omega)$是一个遵从Poisson分布的随机变量.若按它全部可能的不同取值对$\Omega$作分割:

$$\Omega = \bigcup_n \{\omega: \xi(\omega) = n\},$$

求$\xi(\omega)$的熵(也即上述分割的熵).

5. 若$\xi(\omega)$是取值在$[a,b]$上的具有分布密度$p(x)$的随机变量,对$[a,b]$作有限分割

$$a = t_0^n < t_1^n < \cdots < t_n^n = b,$$

令$A_k^n = \{\omega: \xi(\omega) \in [t_{k-1}^n, t_k^n)\}$,则对每一个固定的$n$,$\Omega = \bigcup_{k=1}^{n} A_k^n$是一个$\Omega$的分割$\pi^n$,试求出$H(\pi^n)$,并求出一个尺度变换因子$a_n$,使在$\max_k |t_k^n - t_{k-1}^n| \to 0$时

$$\frac{1}{a_n} H(\pi^n) \to -\int_a^b p(x)\log p(x)\mathrm{d}x.$$

6. 试分别举出一个可数取值与连续取值的随机变量皆具有熵为$+\infty$.

7. 设有$\xi_1(\omega)$与$\xi_2(\omega)$均为仅有有限个不同取值的随机变量,又设$\pi_1$与$\pi_2$分别是按它们各自不同取值所得的$\Omega$的分割(参见4题),试求出$H(\pi_2|\pi_1)$.

8. 设在含有$m$个红球与$n$个白球的一个袋中连续两次抽取一个球,第一次抽得的结果应对应一个分割$\pi_1$,第二次抽取的结果对应分割$\pi_2$,试对有放回的抽取与无放回的抽取两种情形分别研究$H(\pi_2|\pi_1)$,$H(\pi_2)$与$H(\pi_1)$之间的关系.

9. 对12个珠,若其中有一个假珠,未知其轻于还是重于真珠,现用无法码天平去秤,问最少要称几次才能找出假珠来?怎样秤法?

10. 试将秤珠问题的例推广到更为一般的情况(尽力推广到最一般的情况,而得到整齐的结果).

11. 试证明命题 7.8.3 与 7.8.4.

12*. 令 $\Omega=[0,1)$，$F$ 是的一切 Lebegue 可测集，令 $\theta:\Omega\to\Omega$，$\omega\to 2\omega(\mathrm{mod}1)$（$2\omega$ 的小数部分）. 试证明对 Lebegue 测度 $\theta$ 是一个保测变换，而且它是遍历的.

13*. 试证明 2 题中的 $\xi$ 是遍历的，并求出

$$\lim_{n\to\infty}\frac{1}{T}\int_0^T \sin(\xi(\omega)+\theta s)\mathrm{d}s$$

与

$$\lim_{n\to\infty}\frac{1}{T}\int_0^T \sin(\xi(\omega)+\theta s)\sin(\xi(\omega)+\theta(s+\tau))\mathrm{d}s$$

（在概率为 1 的意义下的极限）.

# 第八章　Gauss 过程与二阶矩方法

正态分布是概率论中最常见的分布；Gauss 过程也是一种常见的简单随机过程. 对于具有二阶矩的过程（它的任意时刻的随机变量都有有限二阶矩），如果只限于期望、方差与协方差为依据的研究，其实把它们当作 Gauss 过程并无二致.

## §8.1　Gauss 系

### 1. Gauss 系的概念

**定义 8.1**　设 $I$ 是任意一个有限或无限（甚至不可数）的指标集. 随机变量族 $\{\zeta_t(w); t\in I\}$ 称为一个 Gauss 系，如果对 $I$ 中任意有限个：$t_1, t_2, \cdots, t_n$，都使 $\{\zeta_{t_1}(w), \cdots, \zeta_{t_n}(w)\}$ 遵从 $n$ 维正态分布或退化正态分布，用特征函数的语言即对任意的实数 $\lambda_1, \cdots, \lambda_n$，

$$E\left(e^{i\sum_{k=1}^{n}\lambda_k\zeta_{t_k}(w)}\right)=e^{i\sum_{k=1}^{n}\lambda_k\mu_k-\frac{1}{2}\sum_{k,l}\sigma_{kl}\lambda_k\lambda_l}, \tag{8.1}$$

其中 $\mu_k=E\zeta_{t_k}, \sigma_{kl}=E(\zeta_{t_k}-\mu_k)(\zeta_{t_l}-\mu_l)$.

我们要说明的是：对于正态分布，矩阵 $(\sigma_{kl})$ 是可逆的. $(\zeta_{t_1}-\mu_1, \cdots, \zeta_{t_n}-\mu_n)$ 的联合分布密度是

$$\frac{1}{\left[(2\pi)^n\det(\sigma_{kl})\right]^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x_1,\cdots,x_n)(\sigma_{kl})^{-1}\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}\right). \tag{8.2}$$

但是，当 $(\zeta_{t_1}, \cdots, \zeta_{t_n})$ 遵从退化正态分布时，$(\sigma_{kl})$ 就不可逆，即行列式 $\det(\sigma_{kl})=0$. 当 $I=(-\infty, \infty)$ 或 $[0, \infty]$ 时 Gauss 系称为 Gauss 过程.

**例 8.1** 设$\{\xi_n; n\geqslant 1\}$是一列 i.i.d. 的正态分布 $N(\mu,\sigma^2)$的序列，那么它就是一个 Gauss 系.

**例 8.2** 令 $\eta_i=\sum_{k=1}^{n}a_{ik}\xi_k$，其中$\{\xi_n\}$是上例中的 i.i.d. 正态分布序列，那么$\{\eta_n; n=1,2,\cdots\}$就是一个 Gauss 系(证明见以后).若 $a_{ij}=0$，$a_{21}=a_{11}=a$ 那么 $\eta_1=\eta_2=a\xi_1$，于是$(\eta_1,\eta_2)$的联合分布就是退化分布，它没有二元分布密度. 又例如，当 $a_{4i}=0$，$a_{44}=a_{55}=0$，而且 $a_{5i}=a_{3i}+a_{4i}$ $(i=1,2,3)$则$(\eta_3,\eta_4,\eta_5)$也遵从退化正态分布. 又由于$(\eta_3,\eta_4,\eta_5)$虽然不再有概率分布密度，但又是取连续值的随机向量，我们像(8.1)式中那样，采用特征函数来表达它们的联合分布特性就较方便，而且可以使表达方式与非退化正态分布一致.

下面的定理给出了 Gauss 系的定义的另一个等价的说法：

**定理 8.1** $\{\xi_t; t\in I\}$是 Gauss 系的充要条件是它的任意有限个元 $\xi_{t_1},\cdots,\xi_{t_n}$的任意线性组合 $\eta$ 都是一个一维正态随机变量或常数.

**证明** 必要性：考虑 $\eta=\sum_{k=1}^{n}a_k\xi_{t_k}$，在(8.1)式中取 $\lambda_k=a_k\lambda$ $(k=1,2,\cdots,n)$. 于是

$$E(e^{i\lambda\eta})=E\left\{e^{i\sum_{k=1}^{n}\lambda a_k\xi_{t_k}}\right\}=e^{i\sum_{k=1}^{n}\lambda a_k\mu_k-\frac{1}{2}\sum_{k,l}\lambda^2 a_k a_l\sigma_{kl}}\triangleq e^{i\lambda\mu-\frac{1}{2}\lambda^2\sigma^2}.$$

其中

$$\mu=\sum_{k=1}^{n}a_k\mu_k,\quad \sigma^2=\sum_{k,l}a_k a_l\sigma_{kl}$$

(由于协方差矩阵是非负定的，保证 $\sigma^2\geqslant 0$. 当 $\sigma^2=0$ 时我们以概率为 1 地有 $\eta=\mu$).

充分性：从表面看来定理的条件很弱，只需用一维分布来描述，但是由于线性组合是任意的，因而对任意实数组 $\lambda_1,\cdots,\lambda_n$，令

$\eta(\omega) = \sum_{k=1}^{n}\lambda_k\zeta_{t_k}$，记 $\mu_k = E\zeta_{t_k}$，则

$$E\eta = \sum_{k=1}^{n}\lambda_k\mu_k.$$

$$E(\eta - E\eta)^2 = E\Big[\sum_{k=1}^{n}\lambda_k(\zeta_{t_k} - \mu_k)\Big]^2$$

$$= \sum_{k,l=1}^{n}\lambda_k\lambda_l E((\zeta_{t_k} - \mu_k)(\zeta_{t_l} - \mu_l)) = \sum_{k,l=1}^{n}\lambda_k\lambda_l\sigma_{kl}.$$

若 $\eta$ 遵从正态分布，则 $Ee^{i\lambda\eta} = e^{i\lambda E\eta - \frac{1}{2}E(\eta - E\eta)^2\lambda^2}$，取 $\lambda = 1$，即有

$$Ee^{i\sum_{k=1}^{n}\lambda_k\zeta_{t_k}} = e^{i\sum_{k=1}^{n}\lambda_k\mu_k - \frac{1}{2}\sum_{k,l=1}^{n}\lambda_k\lambda_l\sigma_{kl}}.$$

由定义可知$\{\zeta_t: t \geqslant 0\}$是 Gauss 系.

定理 8.1 给出了一个在说法上比较简单的等价定义. 这样，在验证 Gauss 系时只涉及 1-维的正态分布或常数. 特别对退化的情形就很容易说清楚了. 从定理 8.1 很容易得到一个重要的推论：

**推论 1** Gauss 系的任意个由其元素的线性组合组成的随机变量集仍为 Gauss 系.

**证明** 由于线性组合的任意线性组合也仍是一个线性组合，由定理 8.1 可知推论成立.

## 2. Gauss 系的线性组合的分布

设 $\eta_i = \sum_j a_{ij}\zeta_j (i = 1, \cdots, s)$. 记

$$\sigma_{ij}(\vec{\zeta}) = E((\zeta_i - E\zeta_i)(\zeta_j - E\zeta_j)),$$

即

$$\vec{\eta} = \begin{pmatrix}\eta_1\\ \vdots\\ \eta_s\end{pmatrix} = A\begin{pmatrix}\zeta_1\\ \zeta_2\\ \vdots\\ \zeta_n\end{pmatrix} = A\vec{\zeta}$$

$(A=(a_{ij})_{s\times n})$，于是

$$E\vec{\eta}=\begin{pmatrix}E\eta_1\\E\eta_2\\\vdots\\E\eta_s\end{pmatrix}=A\begin{pmatrix}E\zeta_1\\E\zeta_2\\\vdots\\E\zeta_n\end{pmatrix}=AE\vec{\zeta}.$$

$$\begin{aligned}\sigma_{ij}(\vec{\eta})&\triangleq E((\eta_i-E\eta_i)(\eta_j-E\eta_j))\\&=E[(\vec{\eta}-E\vec{\eta})(\vec{\eta}-E\vec{\eta})^T]_{ij}\\&=E[(A\vec{\zeta}-AE\vec{\zeta})(A\vec{\zeta}-AE\vec{\zeta})^T]_{ij}\\&=E[A(\vec{\zeta}-E\vec{\zeta})(\vec{\zeta}-E\vec{\zeta})^TA^T]_{ij}\\&=[AE((\vec{\zeta}-E\vec{\zeta})(\vec{\zeta}-E\vec{\zeta})^T)A^T]\end{aligned}$$

（这里$[\ ]_{ij}$表示括号内矩阵的第 $i$ 行，第 $j$ 列的那个元素），所以

$$\sum(\vec{\eta})\triangleq(\sigma_{ij}(\eta))=A(\sigma_{ij}(\vec{\zeta}))A^T=A\sum(\vec{\zeta})A^T.\quad(8.3)$$

从(8.3)看出，当 $s>n$ 时$(\eta_1,\cdots,\eta_s)$必定退化；而当 $n>s$ 时，若$(\zeta_1,\cdots,\zeta_n)$非退化，则$(\eta_1,\cdots,\eta_s)$非退化当且仅当 $A$ 的秩是 $s$；当 $n=s$ 时，$(\eta_1,\cdots,\eta_n)$非退化当且仅当$(\zeta_1,\cdots,\zeta_n)$非退化，而且 $A$ 是满秩的.

### 3. Gauss 系的线性闭包

若 $\zeta=\{\zeta_t;\ t\in I\}$是一个 Gauss 系，从其全部可能的线性组合得到的随机变量族仍是 Gauss 系. 此外，下面的定理还告诉我们：若一个 Gauss 系的元素列在概率意义下有极限，则此极限加入此 Gauss 系后仍然是 Gauss 系. 更有甚者，一个 Gauss 系的全部可能的在依概率意义下元素列的极限仍组成 Gauss 系. 由于 Gauss 系的一个元素既可看作是一个线性组合，又可看成是其元素列的极限，因而下面的定理说明 $\zeta$ 的元素的全部线性组合的随机变量族是一个包含原 Gauss 系的扩大的 Gauss 系，记成 $L(\zeta)$；而 $L(\zeta)$中的全部可能的元素列依概率收敛的意义下的极限随机变量族则是

一个更大的 Gauss 系，记成 $L(\xi)$，称之为 $\xi$ 的线性闭包. 它的每个元素都是 $\xi$ 中元素线性组合的极限，因而可看成 $\xi$ 的某个“函数”(更确切地应称为“泛函”)$\varphi(\xi)$.

**定理 8.2** 设 $\xi=\{\xi_t;\ t\in T\}$ 是 Gauss 系，又若 $t_n\in T$，$\xi_{t_n}\xrightarrow{P}\xi_0$（依概率收敛），则 $E|\xi_{t_n}-\xi_0|^2\to 0\ (n\to\infty)$，而且 $\xi_0$ 也遵从 Gauss 分布（正态或退化正态分布）.

**证明** 因为 $\xi_{t_n}\xrightarrow{P}\xi_0$，所以对于 $\forall\varepsilon>0$，

$$P(|\xi_{t_i}-\xi_{t_j}|>\varepsilon)\to 0\quad (i,j\to\infty). \tag{8.4}$$

记 $m_{ij}=E\xi_{t_i}-E\xi_{t_j}$，$\sigma_{ij}^2=\mathrm{Var}(\xi_{t_i}-\xi_{t_j})$，

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{(2\pi)^{\frac{1}{2}}}e^{-\frac{y^2}{2}}\mathrm{d}y.$$

于是(8.4)变成当 $i,j\to\infty$ 时的下述极限式：

$$\frac{1}{\sqrt{2\pi}\,\sigma_{ij}}\int_{|x|>\varepsilon}\exp\left\{-\frac{1}{2}\left(\frac{x-m_{ij}}{\sigma_{ij}}\right)^2\right\}\mathrm{d}x\to 0.$$

但是上式左边等于

$$\frac{1}{\sqrt{2\pi}}\int_{|\sigma_{ij}y+m_{ij}|>\varepsilon}e^{-\frac{y^2}{2}}\mathrm{d}y$$

$$=\left[1-\Phi\left(\frac{-m_{ij}+\varepsilon}{\sigma_{ij}}\right)\right]+\Phi\left(-\frac{m_{ij}+\varepsilon}{\sigma_{ij}}\right).$$

因此(8.4)蕴含：存在 $N$，只要 $i,j>N$ 上式就小于 $1-\Phi(1)$ $(=\Phi(-1))$. 由此我们有

$$\Phi\left(-\frac{m_{ij}+\varepsilon}{\sigma_{ij}}\right)<\Phi(-1),\quad 1-\Phi\left(\frac{-m_{ij}+\varepsilon}{\sigma_{ij}}\right)<1-\Phi(1),$$

从而

$$\frac{m_{ij}+\varepsilon}{\sigma_{ij}}>1,\quad \frac{-m_{ij}+\varepsilon}{\sigma_{ij}}>1\quad (i,j>N\ 时).$$

即 $\sigma_{ij}\pm m_{ij}<\varepsilon$. 把两式相加就得到 $\sigma_{ij}<\varepsilon\ (i,j>N)$. 另一方面还可得到 $\pm m_{ij}<\varepsilon$，即 $|m_{ij}|<\varepsilon\ (i,j>N)$. 可见

$$\sigma_{ij}^2 + m_{ij}^2 \to 0 \quad (i, j \to \infty).$$

所以 $E(\zeta_{t_i} - \zeta_{t_j})^2 = \sigma_{ij}^2 + m_{ij}^2 \to 0 (i, j \to \infty)$. 于是我们得到 $\zeta_{t_j} \xrightarrow{L^2}$ 某个随机变量，而且它必须是 $\zeta_0$，即有

$$E(\zeta_{t_j} - \zeta_0)^2 \to 0 \quad (j \to \infty),$$

由此还得到 $E|\zeta_{t_j} - \zeta_0| \to 0$. 从而

$$E\zeta_{t_j} \to E\zeta_0, E\zeta_{t_j}^2 \to E\zeta_0^2, \quad \mathrm{Var}\zeta_{t_j} \to \mathrm{Var}\zeta_0. \qquad (8.5)$$

进而

$$Ee^{i\lambda\zeta_0} = \lim_{j\to\infty} Ee^{i\lambda\zeta_{t_j}} = \lim_{j\to\infty} e^{i\lambda E\zeta_{t_j} - \frac{1}{2}\lambda^2 \mathrm{Var}\zeta_{t_j}} = e^{i\lambda E\zeta_0 - \frac{1}{2}\lambda^2 \mathrm{Var}\zeta_0}.$$

可见 $\zeta_0$ 仍服从 Gauss 分布.

**推论 1** 设 $\{\zeta_t : t \in I\}$ 是 Gauss 系. 若对 $t_n^{(k)} \in I (k \leqslant m)$ 有 $\zeta_{t_n}^{(k)} \xrightarrow{P} \eta_k (n \to \infty)$，则 $\{\zeta_{t_n}^{(k)} : t \in I\} \cup \{\eta_k : k \leqslant m\}$ 也是一个 Gauss 系.

**证明** 当 $n \to +\infty$ 时，有

$$E(\zeta_{t_n}^{(k)} - \eta_k), \quad E(\zeta_{t_n}^{(k)} - \eta_k)^2 \to 0,$$

可见

$$\mu_k = E\eta_k = \lim_{n\to\infty} E\zeta_{t_n}^{(k)},$$

$$\sigma_{kk} = E(\eta_k - E\eta_k)^2 = \lim_{n\to\infty} E(\zeta_n^{(k)} - E\zeta_n^{(k)})^2,$$

$$\sigma_{kl} = E(\eta_k - E\eta_k)(\eta_l - E\eta_l) = \lim_{n\to\infty} \sigma_{kl}^{(n)},$$

其中 $\sigma_{kl}^{(n)} = E(\zeta_n^{(k)} - E\zeta_n^{(k)})(\zeta_l^{(k)} - E\zeta_l^{(k)})$，于是

$$Ee^{i\sum_k \lambda_k \eta_k} = \lim_{n\to\infty} E(e^{i\sum_k \lambda_k \zeta_{t_n}^{(k)}}) = \lim_{n\to\infty} e^{i\sum_k \lambda_k E\zeta_{t_n}^{(k)} - \frac{1}{2}\sum_{k,l}\lambda_k\lambda_l\sigma_{kl}^{(n)}}$$

$$= e^{i\sum_k \lambda_k \mu_k - \frac{1}{2}\sum_{k,l}\lambda_k\lambda_l\sigma_{kl}},$$

即 $(\eta_1, \cdots, \eta_m)$ 遵从 Gauss 分布.

**定义 8.2** 两个 Gauss 系 $\zeta = \{\zeta_t : t \in I\}$，$\eta = \{\eta_s : s \in J\}$ 称为独立的，如果对任意 $n, m$ 及 $t_1, \cdots, t_n \in I, s_1, \cdots, s_m \in J$，随机向量 $(\zeta_{t_1}, \cdots, \zeta_{t_n})$ 恒与随机向量 $(\eta_{s_1}, \cdots, \eta_{s_m})$ 独立.

由 Gauss 系的定义(8.1)以及独立性用特征函数的描述，我们立刻就得到：

$\xi$ 与 $\eta$ 独立当且仅当对任意 $\forall t\in I$ 及 $s\in J$，恒有

$E[(\xi_t - E\xi_t)(\eta_s - E\eta_s)] = 0$ （即 $\xi_t$ 与 $\eta_s$ 不相关）.

一个 Gauss 系在依概率收敛意义下的线性闭包是一个 Gauss 系，又是一个线性空间，我们可以在这个线性空间上定义内积

$$(\xi,\eta) = E(\xi - E\xi)(\eta - E\eta) \tag{8.6}$$

而形成一个内积空间.

**注**：即使 $I$ 是无限集，这个内积空间总是完备的，它是一个 Hilbert 空间. 通常，为了方便起见，我们都事先从随机变量中扣除其数学期望，然后再对它们进行处理. 因此，这就可以设线性闭包中一切随机变量都是数学期望为 0 的. 这时内积就成为

$$(\xi,\eta) = E(\xi\eta). \tag{8.6'}$$

对于 Gauss 系来说，若 $\xi_{t_1},\cdots,\xi_{t_n}$ 与 $\eta$ 在上述内积下**正交**，即

$$(\xi_{t_k},\eta) = 0 \quad (k = 1,2,\cdots,n),$$

那么对任意实数 $a_1,\cdots,a_n$ 都有

$$\left(\sum_{k=1}^{n} a_k\xi_{t_k},\eta\right) = \sum_{k=1}^{n} a_k(\xi_{t_k},\eta) = 0,$$

也就是说 $\eta$ 与 $\xi$ 中每一个都正交，就意味着它与 $\bar{L}(\xi)$ 中每一个随机变量都正交，也就是它与 $\bar{L}(\xi)$ 独立.

### 4. Gauss 系的投影

在实际问题中，常常有些随机变量不能被直接观测到，而只能通过其他有关的可观测随机变量族去估计它们. 通过可观测的随机变量怎样去估计最为近似呢？下面的命题给出了一个最优估计的判别准则.

设随机变量族 $\varsigma = \{\xi_t : t\in I\}$ 是可观测的，$\eta$ 是一个二阶矩有限的随机变量，不能被直接观测到. 现要从 $\varsigma$ 去估计 $\eta$，就是要找一个随机变量 $\varphi(\varsigma)$（它是 $\varsigma$ 的函数）去近似代替 $\eta$. 我们以均方误

差最小作为最优的标准. 那么 $\varphi(\zeta)$ 是 $\eta$ 的最优估计的含义即是对任意的 $\psi(\zeta)$，只要它也是随机变量，就必须有

$$E(\eta-\varphi(\zeta))^2\leqslant E(\eta-\psi(\zeta))^2. \tag{8.7}$$

**命题 8.2** $\varphi(\zeta)$ 是 $\eta$ 的最优估计当且仅当对任何 $\psi(\zeta)$，只要它是随机变量，且 $E|\psi(\zeta)|^2<+\infty$，就有

$$E[(\eta-\varphi(\zeta))\psi(\zeta)]=0. \tag{8.8}$$

**注**：在命题中对什么样的 $\varphi(\cdot)$ 满足(8.7)的条件以及为什么它有二阶矩并未交代清楚，因为缺少测度论语言，无法交代清. 其实满足(8.7)的 $\varphi(\zeta)$ 就是所谓 $\eta$ 对 $\zeta$ 的"条件期望".

**证明** 若对一切随机变量 $\psi(\zeta)$，(8.8)成立，则

$$\begin{aligned}E(\eta-\psi(\zeta))^2&=E(\eta-\varphi(\zeta)+\varphi(\zeta)-\psi(\zeta))^2\\&=E(\eta-\varphi(\zeta))^2+2E(\eta-\varphi(\zeta))(\varphi(\zeta)-\psi(\zeta))\\&\quad+E(\varphi(\zeta)-\psi(\zeta))^2\\&=E(\eta-\varphi(\zeta))^2+E(\psi(\zeta)-\varphi(\zeta))^2\\&\geqslant E(\eta-\varphi(\zeta))^2.\end{aligned}$$

反之，若 $\varphi(\zeta)$ 是最优估计，则对任意实数 $\lambda$ 有

$$E(\eta-\varphi(\zeta))^2\leqslant E(\eta-\varphi(\zeta)+\lambda\psi(\zeta))^2.$$

从而

$$\lambda^2E(\psi(\zeta))^2+2\lambda E[(\eta-\varphi(\zeta))\psi(\zeta)]\geqslant 0.$$

因此必须有

$$E[(\eta-\varphi(\zeta))\psi(\zeta)]=0.$$

对于一般随机变量族 $\zeta$ 与 $\eta$，要想验证 $\varphi$ 满足(8.8)是非常困难的，但是当 $\zeta\cup\{\eta\}$ 组成 Gauss 系时验证(8.8)可以大大地简化.

为了方便起见，不失一般性地我们设 $E\zeta_t=0\ (\forall t\in I)$ 与 $E\eta$ 等于零的情形(否则先将期望分别减去后再行处理).

**命题 8.2′** 设 $\zeta\triangleq\{\zeta_t,t\in I\}$ 与 $\eta$ 一起组成一个均值全为 0 的 Gauss 系，则 $\eta$ 在 $\zeta$ 的线性闭包 $\bar{L}(\zeta)$ 中的投影 $\hat{\eta}$(即 $\hat{\eta}$ 应对 $\forall t\in I$，都有 $E[(\eta-\hat{\eta})\zeta_t]=0$) 就是由 $\zeta$ 对 $\eta$ 的最优估计.

**证明** 由$\hat{\eta}$是$\eta$在$\overline{L}(\zeta)$中的投影可见能有表示$\hat{\eta}=\varphi(\zeta)$，而且对任一个$t\in I$都成立$E((\eta-\hat{\eta})\zeta_t)=0$. 由Gauss系的特点可以推出$\eta-\hat{\eta}$与$\zeta$独立，也就与任一形如$\psi(\zeta)$的随机变量独立，因此有

$$E[(\eta-\hat{\eta})\psi(\zeta)]=E(\eta-\hat{\eta})E\psi(\zeta)=0.$$

再由命题8.1，就得到$\hat{\eta}$是由$\zeta$可作的最优估计.

**例8.3**（滤波） 设有信号（随机序列）$S=\{S_n: n=1,2,\cdots,m\}$，又有与之独立的噪声$N=\{N_n;\ n=1,2,\cdots,m\}$的干扰，它们都分别是期望为0的Gauss系. 已知$R_{n,k}=E(S_nS_k)$（$n,k=1,\cdots,m$）及$r_{n,k}=E(N_nN_k)$. 试求出由$\{X_n=S_n+N_n;\ n=1,2,\cdots,m\}$去估计$S_t$（$t=1,2,\cdots,m$）的最佳估计的公式（通常人们将这里的问题称为滤波，意思是通过一个“滤波”运算，将噪声“滤去”）.

**解** 将$S_t$的估计记为$\hat{S}_t$，它应是$S_t$在$\{X_t;\ t=1,2,\cdots,m\}$中的投影，因而$\hat{S}_t=\sum\limits_{k=1}^{N}a_kX_k$. 而且满足下列方程组

$$(S_t-\hat{S}_t,\ X_n)=0\ (n=1,\cdots,m),$$

即对任意$n\leqslant m$有

$$E(S_tX_n)=E\{\sum_{k=1}^{m}a_kX_kX_n\}=\sum_{k=1}^{m}a_kE(X_kX_n). \qquad (8.9)$$

由上式中$\{S_n\}$与$\{N_n\}$的独立性，我们有

$$\text{左}=E[S_t(S_n+N_n)]=ES_tEN_n+R_{t,n}=R_{t,n},$$

$$\text{右}=\sum_{k=1}^{m}a_kE[(S_k+N_k)(S_n+N_n)]=\sum_{k=1}^{m}a_k(R_{k,n}+r_{k,n}).$$

于是(8.9)就变为线性方程组

$$\sum_{k=1}^{m}(R_{k,n}+r_{k,n})a_k=R_{t,n}\quad(n=1,2,\cdots,m). \qquad (8.10)$$

若矩阵$\bar{R}\triangleq(R_{k,n}+r_{k,n})_{k,n=1,2,\cdots,m}$可逆，则

$$\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} = \widetilde{R}^{-1} \begin{pmatrix} R_{t,1} \\ \vdots \\ R_{t,m} \end{pmatrix}.$$

当 $R$ 没有逆时，我们可以用伪逆去给出解，这里略去详情．特别当时齐情形，即 $R_{s,t}=R(|t-s|)$，$r_{s,t}=r(|t-s|)$（只与差有关）时，上面的(8.10)解法可以纳入宽平稳列的框架来解决(参见 8.3).

### 5. 复 Gauss 系

**定义 8.3** 复随机变量系 $\xi \triangleq \{\zeta_t + i\eta_t;\ t\in T\}$ 称为复 Gauss 系，如果 $\zeta,\eta$ 是两个相互独立且 $(\zeta - E\zeta)$ 与 $(\eta - E\eta)$ 相同分布(指所有相应的有限维随机变量组 $(\zeta_{t_1}-E\zeta_{t_1},\cdots,\zeta_{t_n}-E\zeta_{t_n})$ 与 $(\eta_{t_1}-E\eta_{t_1},\cdots,\eta_{t_n}-E\eta_{t_n})$ 具有相同的分布).

**注**：由于技术上的原因，复 Gauss 系的要求比实 Gauss 系高得多，这个定义的缺陷是实 Gauss 系并不能看成虚部为 0 的复 Gauss 系．然而这个定义的好处是使实 Gauss 系的几乎所有结果都在复 Gauss 系平行地成立，唯一的改变是协方差函数在复 Gauss 系应定义为

$$R(t,s) \triangleq E[(\xi_t - E\xi_t)\overline{(\xi_s - E\xi_s)}]$$

(其中 $\bar{\ }$ 代表求复数的共轭运算)．由于在 Fourier 分析中常出现乘子 $e^{i\lambda}$，于是用上述复过程去研究实过程，常常非常方便．

如果 $E(\zeta + i\eta)\overline{(\zeta' + i\eta')} = 0$，则称 $\zeta + i\eta$ 与 $\zeta' + i\eta'$ 正交．

## §8.2 具有二阶矩的过程

### 1. 宽平稳列概述

**定义 8.4** 随机变量列 $\{\zeta_n : n\geqslant 0\}$ 称为**宽平稳的**，如果 $E\zeta_n$，$E\zeta_n\zeta_{n+m}$ 均不依赖 $n$．记

$$B(m) = E(\zeta_n\zeta_{n+m}),\quad R(m) = E((\zeta_n - E\zeta_n)(\zeta_{n+m} - E\zeta_{n+m})). \tag{8.11}$$

它们分别称为$\zeta_n$的**相关序列**及**协方差序列**. 宽平稳列一般也称为宽平稳时间序列. 实际中的一个重要问题是：用$\{\zeta_n: n\geqslant 0\}$的一段观测值估计协方差序列$R(m)$及期望$E\zeta_1$. 下面记$a=E\zeta_1$. 对此我们有(但在本书中不予以证明)

**定理 8.3** 设$\zeta_n$为宽平稳列,$E\zeta_n=\mu$,如果其协方差列$R(m)$满足:$\exists\alpha>0$使

$$\frac{1}{n^2}\sum_{i,j=1}^{n}R(i-j)=\frac{1}{n}\sum_{k=-(n-1)}^{n-1}\left(1-\frac{|k|}{n}\right)R(k)=O\left(\frac{1}{n^\alpha}\right), \tag{8.12}$$

则

$$P\left\{\frac{1}{n}\sum_{k=1}^{n}\zeta_k\xrightarrow{n\to\infty}E\zeta_1\right\}=1.$$

特别,当$\sum_k|R(k)|<\infty$时,定理条件(8.12)满足.

**意义** 用一段现实观测$\zeta_1,\cdots,\zeta_n$的平均去估计$E\zeta_1$,这个定理的结论说明这种估计是合理的(在数理统计上叫做强相合的).

**定理 8.4** 设$\zeta_n$为宽平稳 Gauss 列(此时$\zeta_n$必是平稳列,故是平稳 Gauss 列),且$E\zeta_n=0$,又若

$$\frac{1}{n}\sum_{k=1}^{n}|R(k)|^2\to 0, \tag{8.13}$$

则

$$P\left\{\frac{1}{n}\sum_{k=1}^{n}\zeta_k\zeta_{k+m}\to R(m)\right\}=1.$$

又若$|R(n)|\leqslant\frac{c}{n^\alpha}(\alpha>0)$,则(8.13)满足.

**意义** 此时$R(m)$可用$\zeta_1\zeta_{1+m},\cdots,\zeta_n\zeta_{n+m}$的平均值来估计,并且此估计也是强相合的.

在研究宽平稳过程时,由于常用频谱分析(Fourier 分析),故常用复的随机序列$\zeta_n=\eta_n+i\xi_n$,这时相关函数及协方差函数的定义如第 5 段复 Gauss 系那样:相关函数$B(m)\triangleq E(\zeta_n\bar{\zeta}_{n+m})$. 相关函数具有下述重要性质

(B.1) $B(0)\geqslant 0$;

(B.2) $|B(m)|\leqslant B(0)$ (这是 Schwarz 不等式);

(B.3) $B(-m)=\overline{B}(m)$;

(B.4) 对$\forall m_1,\cdots,m_k$, $(\forall k)$,矩阵$(B(m_i-m_j))_{i,j\leqslant n}\geqslant 0$(非负定 Hermit 矩阵).

**2. 向量值随机过程**

**定义 8.5** 依赖时间参数 $t\in T$ 的随机向量族$\vec{\xi}\triangleq\{\vec{\xi}_t,t\in T\}$称为向量值随机过程,特别若 $T$ 只取正整数时刻$\vec{\xi}$称为向量值随机列. 在不会出现混淆时,我们常把"向量值"三字省去. $\vec{m}(t)\triangleq E\vec{\xi}_t(t\in T)$称为$\vec{\xi}$的均值向量值函数. $R(t,s)\triangleq E(\vec{\xi}_t\,\vec{\xi}_s^T)$称为$\vec{\xi}_t$ 的协方差矩阵值函数,其中$\vec{\xi}_t^T$ 为$\vec{\xi}_t$ 的转置向量.

## §8.3 ARMA 模型

**1. 一般概念**

在第六章中我们曾讨论过作为马氏过程应用的特例的 AR 模型与 ARMA 模型. 这里,我们从二阶矩过程的观点,简单地作一个介绍.

**定义 8.6** 若 $\varepsilon_n$ 为不相关随机变量列且 $E\varepsilon_n=0,E\varepsilon_n\varepsilon_m=\delta_{nm}$,多项式 $\alpha(z)\triangleq 1+\alpha_1 z+\cdots+\alpha_p z^p$ 在$|z|\leqslant 1$ 无根,则称满足如下差分方程

$$\xi_n+\alpha_1\xi_{n-1}+\cdots+\alpha_p\xi_{n-p}=\beta_0\varepsilon_n\quad(\beta_0>0),\qquad(8.15)$$

且 $E\xi_n=0$ 的宽平稳列 $\xi_n$ 为 $p$ 阶**自回归模型**(Auto Regressive). 简称为 AR($p$)**模型**,也称 $p$ **阶宽马氏模型**(注意,第六章§2 例 7 中的 $\mu^k-\lambda_1\mu^{k-1}\cdots-\lambda_k$ 的根,即这里的 $\alpha(1/z)=0$ 的根,$k=p,\alpha_i=-\lambda_i$).

**定义 8.7** 若 $\varepsilon_n$ 同定义 8.2、多项式

$$\beta(z) \triangleq \beta_0 + \beta_1 z + \cdots + \beta_q z^q$$

在 $|z|\leqslant 1$ 无根,则称如下的 $\xi_n$ 为 $q$ 阶**滑动平均模型**(Moving Average),简称 MA($q$)**模型**:

$$\xi_n \triangleq \beta_0\varepsilon_n + \beta_1\varepsilon_{n-1} + \cdots + \beta_q\varepsilon_{n-q}. \tag{8.16}$$

要求在 $|z|\leqslant 1$ 无根是需要的,我们可举如下反例:

**例 8.4** $\xi_n \triangleq \varepsilon_n + 2\xi_{n-1} = \varepsilon_n + 2\varepsilon_{n-1} + \cdots + 2^{n-1}\varepsilon_1 + 2^n\xi_0$ 对初值 $\xi_0$ 非常敏感,因而不可能是稳定的系统.

把这两者统一起来的模型为:

**定义 8.8** 若 $\varepsilon_n$ 同定义 8.6,又设在定义 8.6 及定义 8.7 中多项式 $\alpha(z)$,$\beta(z)$ 在 $|z|\leqslant 1$ 无根,宽平稳列 $\xi_n$ 满足:$E\xi_n=0$ 及

$$\xi_n + \alpha_1\xi_{n-1} + \cdots + \alpha_p\xi_{n-p} = \beta_0\varepsilon_n + \beta_1\varepsilon_{n-1} + \cdots + \beta_q\varepsilon_{n-q}. \tag{8.17}$$

则称 $\xi_n$ 为 $(p,q)$ 阶**自回归滑动平均模型**,简称 ARMA($p$,$q$)**模型**.

**例 8.5** 在军事用途的散射通讯中,由于大气介质的变化而出现电波的涨落现象. P. Monsen 在 1971 年提出广带反馈的均衡器以克服这现象造成的失误,其原理图如下:

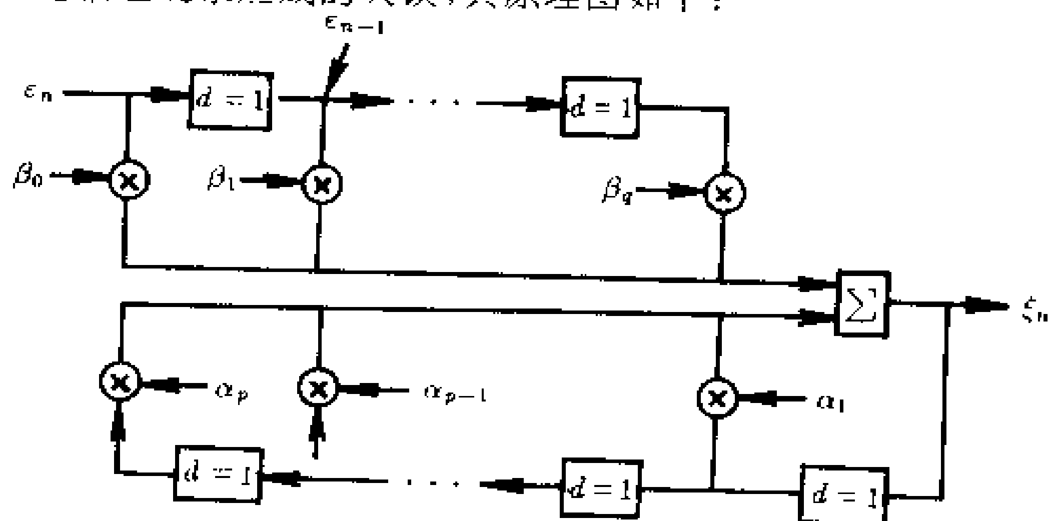



图 8.1

其中 $\{\varepsilon_n\}$ 是收到的信号,通常是带信息的伪随机列,在统计学上伪随机列的性质非常相似于i.i.d.列,所以在理论上把它看成i.i.d.列.⊗表示乘法器.$\Sigma$ 表示加法器,$d=1$ 表示一个单位的时间延迟,而 $\xi_n$ 是通过整个带反馈的均衡器加工处理后所得的信号.只

要对加权系数 $\alpha_i,\beta_j(i\leqslant p,j\leqslant q)$ 加以适当的限制，使 $A(z),B(z)$ 在 $|z|\leqslant 1$ 无根，则 $\xi_n$ 就是 ARMA$(p,q)$ 模型.

**2. AR 模型**

首先的问题是：如何判别一个期望为 0 的宽平稳列是 AR$(p)$，或者是 MA$(q)$，或者是 ARMA$(p,q)$，或者什么都不是. 当然判别的条件应加在协方差列（此时即相关序列）上，为此需要引进一个概念.

**定义 8.9** 设 $\{\xi_n:n\geqslant 0\}$ 是宽平稳列，定义 $\alpha_j^{(k)}$ 为

$$E\left|\xi_n-\sum_{j=1}^{k}\alpha_j^{(k)}\xi_{n-j}\right|^2=\inf_{c_1,\cdots,c_k}E\left|\xi_n-\sum_{j=1}^{k}c_j\xi_{n-j}\right|^2. \tag{8.18}$$

最后一个系数 $\alpha_k^{(k)}$ 称为**第 $k$ 步偏相关数**，而 $\{\alpha_k^{(k)}:k\geqslant 1\}$ 称为 $\{\xi_n:n\geqslant 0\}$ 的**偏相关系数列.**

AR 模型可以用偏相关系数列刻画.

**定理 8.5** 如果宽平稳列 $\xi_n$ 满足 $E\xi_n=0$，且∃ 函数 $f\geqslant 0$，使

$$\int\log f(\lambda)\mathrm{d}\lambda>-\infty,$$

且 $R(n)=\int \mathrm{e}^{\mathrm{i}n\lambda}f(\lambda)\mathrm{d}\lambda$（在处理实际问题时总认为此条件是满足的），那么 $\xi_n$ 是 AR$(p)$ 的充要条件是第 $p$ 个偏相关数非 0，而从第 $p+1$ 个往后偏相关系数均为 0，即

$$\alpha_p^{(p)}\neq 0,\ \alpha_k^{(k)}=0\ \ (k>p). \tag{8.20}$$

如何求偏相关系数列呢？注意到 $\alpha_j^{(k)}(j\leqslant k)$ 是极值问题(8.18)的解. 记(8.18)等号右方的量为 $F(\alpha_1^{(k)},\cdots,\alpha_k^{(k)})$，于是 $(\alpha_1^{(k)},\cdots,\alpha_k^{(k)})$ 满足 $\dfrac{\partial F}{\partial\alpha_j^{(k)}}=0\ (j\leqslant k)$，也就是

$$\begin{pmatrix} R(0) & R(1) & \cdots & R(k-1)\\ R(1) & \cdots & \cdots & \cdots\cdots\\ \vdots & \ddots & \ddots & \vdots\\ & & & R(1)\\ R(k-1) & \cdots & R(1) & R(0)\end{pmatrix}\begin{pmatrix}\alpha_1^{(k)}\\ \vdots\\ \alpha_k^{(k)}\end{pmatrix}=\begin{pmatrix}R(1)\\ \vdots\\ R(k)\end{pmatrix}. \tag{8.21}$$

解这个方程组可得 $\alpha_k^{(k)}$. 然而在定理 8.5 条件下对不同的 $k$ 我们有求 $\{\alpha_j^{(k)}\}$ 的递推算法:

$$
\begin{cases}
\alpha_1^{(1)} = R(1)/R(0), \\
\alpha_2^{(2)} = (R(2) - R(1)\alpha_1^{(1)})(R(0) - R(1)\alpha_1^{(1)})^{-1}, \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\alpha_{k+1}^{(k+1)} = \left(R(k+1) - \sum_{j=1}^{k} R(k+1-j)\alpha_j^{(k)}\right)\left(R(0) - \sum_{j=1}^{k} R(j)\alpha_j^{(k)}\right)^{-1}, \\
\alpha_j^{(k+1)} = \alpha_j^{(k)} - \alpha_{k+1}^{(k+1)}\alpha_{k+1-j}^{(k)} \quad (j \leqslant k).
\end{cases}
\tag{8.22}
$$

这个递推公式的推导并不复杂，但是本书着重于应用，故把它略去. 判定了 AR($p$) 以后，就需用相关函数确定未知参数，这就是:

**定理 8.6** 若 $\xi_n$ 为 AR($p$)，且 $E\xi_n = 0$，则

$$
\begin{pmatrix}
R(0) & R(1) & \cdots & R(p-1) \\
R(1) & \cdots & \cdots & \cdots\cdots \\
\vdots & \ddots & \ddots & \vdots \\
 & & & R(1) \\
R(p-1) & \cdots & R(1) & R(0)
\end{pmatrix}
\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix}
=
\begin{pmatrix} R(1) \\ \vdots \\ R(k) \end{pmatrix},
\tag{8.23}
$$

$$
\beta_0^2 = 1 + \sum_{k=1}^{p} R(k)\alpha_k. \tag{8.24}
$$

### 3. MA 模型

MA($q$) 的判别，就比 AR($p$) 的判别容易多了. 我们有

**定理 8.7** 如果宽平稳列 $\xi_n$ 满足: $E\xi_n = 0$，且∃函数 $f \geqslant 0$，使

$$
R(n) = \int e^{in\lambda} f(\lambda)\,d\lambda
$$

(在实际问题处理中，也总认为这个条件是满足的)，那么 $\xi_n$ 是 MA($q$) 的充分必要条件为

$$
R(q) \neq 0, \quad R(k) = 0 \quad (|k| > q).
$$

在条件满足时，未知参数 $\beta_1, \cdots, \beta_q$ 可以由二次方程组

$$
R(m) = \sum_{k=m}^{q} \beta_k \beta_{k-m} \quad (0 \leqslant m \leqslant q) \tag{8.25}
$$

解得(但需要满足约束条件 $\beta_0>0$). 此方程组可以用第六章 § 6.3 中的方法求近似解(求差的平方和的最小值).

### 4. ARMR 模型

对于 ARMR$(p,q)$的判别,则有

**定理 8.8** 在定理 8.7 同样的条件下,$\xi_n$ 对最小的$(p,q)$是 ARMR$(p,q)$的充要条件为:∃ 根皆在单位园外的 $p$ 阶多项式

$$\alpha(z)=\sum_{j=0}^{p}\alpha_k z^k \quad (\alpha_0=1),$$

使得

$$\sum_{k=0}^{p}\alpha_k R(m-k)=0 \quad (m\geqslant q)\,. \tag{8.26}$$

在条件满足时,未知系数 $\alpha_k$ 们即由 $\alpha(z)$决定,而 $\beta_i$ 们则由下列非线性方程组决定:

$$\sum_{k=0}^{p}\alpha_k R(m-k)=\sum_{j=m}^{q}\beta_j\beta_{j-m} \quad (0\leqslant m\leqslant q). \tag{8.27}$$

## § 8.4 AR 模型的线性预测问题与 Kalman-Bucy 滤波

### 1. 预测问题

时间序列的一个重要问题是预测. 在卫星轨道测定、控制及反弹道武器等问题的研究中,需要由实测到的有限个时刻目标所在位置来快速地预估它在下一些时刻的位置. 也就是说,在已知时间序列的某一段样本值 $\xi_{n-l},\xi_{n-l+1},\cdots,\xi_n$ 的条件下,要预估 $\xi_{n+1},\cdots,\xi_{n+k}$. 由 § 8.1 命题 8.2 我们还可以看到,当$\{\xi_n\}$是 Gauss 系时,最优预测一定是一个线性预测.

最简单的是在线性函数 $\gamma_0\xi_n+\gamma_1\xi_{n-1}+\cdots+\gamma_l\xi_{n-l}$中寻最优的系数$(\gamma_0,\gamma_1,\cdots,\gamma_l)$,使 $E|\gamma_0\xi_n+\cdots+\gamma_l\xi_{n-l}-\xi_{n+1}|^2$ 最小. 这就变成

一个简单的极值问题．此外，对高速飞行器而言，由于速度高，轨道在每个单位时刻中总能用切线方向很好地代替，因而线性预测虽不是最优的，却不失为一种较好的快速简便预测．

对于 AR($p$) 而言，由模型直接可以看出取长度为 $p$ 的一段观测 $\xi_n,\xi_{n-1},\cdots,\xi_{n-p+1}$ 就充分地反映了信息，由于 $\xi_{n+1}+\alpha_1\xi_n+\cdots+\alpha_p\xi_{n-p+1}=\beta_0\varepsilon_{n+1}$ 等式右端不包含 $n$ 前的信息，于是 $\xi_{n+1}$ 的预测值自然地取成

$$\hat{\xi}_{n+1}\triangleq-(\alpha_1\xi_n+\alpha_2\xi_{n-1}+\cdots+\alpha_p\xi_{n-p+1}).\qquad(8.28)$$

而 $\xi_{n+2}$ 的预测值可取成

$$\hat{\xi}_{n+2}\triangleq-(\alpha_1\hat{\xi}_{n+1}+\alpha_2\xi_n+\cdots+\alpha_p\xi_{n-p+2}).$$

如此下去，利用(8.28)就能用 $\{R(n)\}$ 直接作预测．

直观上，ARMA 模型可以用 AR 模型近似．事实上，ARMA 模型的 $R(n)$ 总可写成某个 $f(\lambda)$ 的 Fourier 系数，因而总有 $R(n)\to0$，因此只要取 $p$ 充分大，使 $R(n)\approx0(n\geqslant p)$，使用 AR($p$) 代替原来的模型就足够好了．

当然 ARMA 模型的预测问题也有显式表达，实际上在下一段的 Kalman-Bucy 滤波已经包容了这个模型．

**例 8.6** 某地平均年降雨 540mm，已知其偏差（降雨量－平均降雨）满足 AR(2) 模型

$$\xi_n+0.54\xi_{n-1}-0.3\xi_{n-2}=\varepsilon_n.$$

又已知近五年实际降雨量为 560，470，585，496，576，相应偏差各为 20，－70，45，－44，36，于是下一年的偏差预测值为：$\hat{\xi}_1\triangleq-0.54\xi_0+0.3\xi_{-1}$，即

$$\hat{\xi}_1\triangleq-0.54\times36+0.3\times(-44)=-32.64,$$

再下一年偏差预测为：

$$\hat{\xi}_2\triangleq-0.54\hat{\xi}_1+0.3\xi_0=28.43.$$

第三年偏差预测为：

$$\hat{\xi}_3 \triangleq -0.54\hat{\xi}_2 + 0.3\hat{\xi}_1 = -25.14 .$$

这样三年内的降雨量预测值分别为 507.36,568.43,514.86.

**2. 滤波问题**

1）背景

在观测信号中混有噪声干扰时，就要利用信号本身的规律去滤掉噪声. 这就是滤波问题. §8.1.4 中例 8.3 就是这一类问题.

最简单的信号模型是 AR($p$)模型

$$\xi_n = \sum_{j=1}^{p} \alpha_j \xi_{n-j} + w_n. \tag{8.29}$$

(8.29)称为**状态方程**(为了与文献一致，我们把 §8.3 中的 $\varepsilon_n$ 改写为 $w_n$). 而实测到的数是由 $\{\xi_n : n \geqslant 0\}$ 在空间传播的方式与量测噪声 $v_n$ 决定的. 假定信号通过多个通道使到达各接收点时间并不一致，于是汇总各接收点而接收到的测量值不是信号 $\xi_n$，而是显现信号过去值的不同权重的滑动和再加上噪声：

$$\sum_{j=0}^{p-1} h_j \xi_{n-j} + v_n.$$

我们把它记为 $\eta_n$，即

$$\eta_n = \sum_{j=0}^{p-1} h_j \xi_{n-j} + v_n, \tag{8.30}$$

称为**量测方程**.

(8.29),(8.30)构成了下面将要讲的 Kalman-Bucy 模型的核心，它比早先的 $\eta_n = \xi_n + v_n$ 的模型更切合实际，且包含了后者.

为了数学上与实用上方便，我们还需要把模型改写成矩阵形式，并把数学模型进一步叙述清楚，(8.29)可以写成

$$\begin{pmatrix} \xi_n \\ \xi_{n-1} \\ \vdots \\ \xi_{n-p+1} \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_p \\ 1 & 0 & \cdots & \\ & \ddots & & \\ 0 & & 1 & 0 \end{pmatrix} \begin{pmatrix} \xi_{n-1} \\ \xi_{n-2} \\ \vdots \\ \xi_{n-p} \end{pmatrix} + \begin{pmatrix} w_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{8.29$'$}$$

即可以写成下述矩阵形式

$$\vec{x}_n = F\vec{x}_{n-1} + \Gamma \vec{w}_n \text{（且 } F \text{ 的特征根均在 } |z| \leqslant 1 \text{ 外面）}. \qquad (8.29)''$$

而(8.30)则可以写成

$$\eta_n = (h_1, h_2, \cdots, h_p) \begin{pmatrix} \xi_n \\ \xi_{n-1} \\ \vdots \\ \xi_{n-p+1} \end{pmatrix} + v_n, \qquad (8.30)'$$

即写成下述矩阵形式（$\eta_n$ 改写成 $\vec{y}_n$）：

$$\vec{y}_n = H\vec{x}_{n-1} + \nu_n. \qquad (8.30)''$$

(8.29)″与(8.30)″合称为 **Kalmom-Bucy 模型**. 事实上，模型(8.29)″远比(8.29)为广. 在数学上可以证明如下的定理：

**Gauss 定理** 如果$\{\xi_n\}$为 ARMA$(p,q)$，则存在满足(8.29)″中的模型（适当选取 $F$、$\Gamma$、$\{w_n\}$ 及 $\vec{x}_n$ 的分量），使 $\xi_n$ 恰是 $\vec{x}_n$ 的第一个分量. 因此(8.29)″包含了 ARMA$(p,q)$，所以可以认为 Kalman-Bucy 模型是 ARMA$(p,q)$的自然推广.

再者(8.29)″，(8.30)″也比(8.29)′，(8.30)′更为一般. 在(8.29)″，(8.30)′中 $\omega_n$、$v_n$、$\eta_n$ 都是一维的，$\Gamma$ 为 $p\times 1$，$H$ 为 $1\times p$ 矩阵. 而(8.29)″，(8.30)″将可以讨论更普遍的情形.

2) Kalman-Bucy 滤波模型

方程组

$$\vec{x}_n = F\vec{x}_{n-1} + \Gamma\vec{w}_n, \qquad (8.29)''$$

$$\vec{y}_n = H\vec{x}_{n-1} + \vec{\nu}_n \qquad (8.30)''$$

（其中 $\vec{x}_n$ 是 $p$ 维向量值过程，$\vec{W}_n$ 是 $r$ 维向量值过程，$\vec{y}_n$、$\vec{\nu}_n$ 是 $m$ 维向量值过程，$F: p\times p$，$\Gamma: p\times r$，$H: m\times p$，$\vec{W}_n$、$\vec{\nu}_n$ 均值为 0 的、相互独立的白噪声列（独立同分布列）$E(\vec{W}_n\vec{W}_n^T)=Q$，$E(\vec{\nu}_n\vec{\nu}_n^T)=R$，右上角的 $T$ 表示矩阵的转置运算，这里$\{\vec{x}_n\}$是状态列，代表信号，

$\{\vec{y}_n\}$是对$\{\vec{x}_n\}$的观测列，而$\{\vec{v}_n\}$是观测误差列）. 给出了 Kalman-Bucy 滤波模型，它们分别称为状态方程及量测方程. Kalman-Bucy 滤波就是要从观测列$\vec{y}_n$出发递推地对$\vec{x}_n$作估计，即利用对$\vec{x}_{n-1}$的估计值（记为$\hat{x}_{n-1}$）及$\vec{y}_n$给出对$\vec{x}_n$的估计值$\hat{x}_n$. 它的原则是：若已知$\hat{x}_{n-1}$，则选择矩阵 $A,B$，使方差矩阵

$$\Delta_n \triangleq E[(\vec{x}_n-(A\hat{x}_{n-1}+B\vec{y}_n))(\vec{x}_n-(A\hat{x}_{n-1}+B\vec{y}_n))^T]$$

极小（对称非负矩阵称为 $C_1\geqslant C_2$，如果 $C_1-C_2$ 为非负定），并取

$$\hat{x}_n=A\hat{x}_{n-1}+By_n$$

作为对$\vec{x}_n$的估计，$\{\hat{x}_n\}$称为$\{\vec{x}_n\}$的滤波列. 从$\{\vec{y}_n\}$得到滤波值列$\{\hat{x}_n\}$的装置称为 Kalman-Bucy 滤波器.

注意到 $\Delta_n$ 是 $A$ 与 $B$ 的“二次式”，可以由“配平方”方法求 $\Delta_n$ 的极小，这个做法思想很简单，但是写起来有点累赘，所以我们直接写出结果：

$$\begin{cases}A=(I-BH)F,\\ B=\Pi_nH^T(H\Pi_nH^T+R)^{-1}\quad(\text{记成 }K_n),\end{cases}$$

其中

$$\Pi_n=FP_{n-1}F^T+\Gamma Q\Gamma^*,$$

$$P_n\triangleq E(\vec{x}_n-\hat{x})\ (\vec{x}_n-\hat{x})^T\quad(\text{滤波方差阵}).$$

把 $B$ 改记为 $K_n$，那么

$$\begin{aligned}\hat{x}_n&=(I-K_nH)F\hat{x}_{n-1}+K_n\vec{y}_n\\ &=F\hat{x}_{n-1}+K_n(\vec{y}_n-HF\hat{x}_{n-1}).\end{aligned}$$

$K_n$ 又称为（噪声）**增益系数**，上式右端第二项是量测方程作出的“附加”贡献.

这里 $P_0$ 的取值并不很重要，而$\vec{x}_0$可由我们对系统的认识来给定，有时$\hat{x}_0$可取为 0，而 $P_0$ 可取成初值$\vec{x}_0$的方差阵，因为在

Kalman-Bucy 滤波理论中有一条重要定理：在 $F,H,\Gamma$ 满足一个相当弱的条件(称为可观测条件)下，当 $n\to\infty$ 时 $P_n$ 总有与 $P_0$ 无关的极限. 这说明只要 $n$ 充分大，特殊取的 $P_0$ 所产生的影响是可以消除的. 同时由于 $P_n\to P$, $K_n$ 应趋向某个 $K$. 也就是递推到适当步数后 $K$ 可以不必再变动计算了，这就减少了计算量.

综上看出，Kalman-Bucy 滤波必须先有模型参数 $F,H,\Gamma,Q$, $R$. 从一些"训练"数据 $\vec{x}_1,\cdots,\vec{x}_n,\vec{y}_1,\cdots,\vec{y}_n$ 来估计这些参数，这是模型的优化拟合问题. 一般地它的计算量较大，在实用中，如果用(8.29)′, (8.30)′ 近似地替代，有时也能得到合适的近似.

**注**：$F,H,\Gamma,Q,R$ 可以依赖于 $n$，这时得到的是非时齐的 Kalman-Bucy 滤波器.

进一步还有不少技术性问题，有兴趣的读者可参考：

《离散时间系统的数学方法》，中科院数学所概率组，国防工业出版社，1975.

## §8.5 附　　录

### 1. 条件期望的直观定义

若$(\xi,\eta)$是离散随机变量，记

$$E(\xi|\eta=y_j)=\sum_i x_i P(\xi=x_i|\eta=y_j),$$

它是 $\eta=y_j$ 条件下 $\xi$ 的条件分布的数学期望. 显见它是 $y_j$ 的函数. 不妨将这个函数记成 $\varphi(y_j)$. 那么

$$E(\xi|\eta=y_j)=\varphi(y_j).$$

另一方面，如果$(\xi,\eta)$具有联合分布密度 $f(x,y)$，那么在 $\eta=y$ 条件下，$\xi$ 的条件分布密度为

$$f(x|y)\triangleq\frac{f(x,y)}{f_2(y)},$$

其中 $f_2(y)$ 为 $\eta$ 的边缘密度 $f_2(y)=\int f(x,y)\mathrm{d}x$. $\xi$ 关于这个条件密度的数学期望记成

$$E(\xi|\eta=y)=\int xf(x|y)\mathrm{d}x,$$

它也是 $y$ 的函数.我们也把它记成 $\varphi(y)$,即 $E(\xi|\eta=y)=\varphi(y)$. 当 $\eta$ 不取定值 $y$ 时,$\varphi(\eta)$ 是一个随机变量,于是自然地可以定义:$E(\xi|\eta)\triangleq\varphi(\eta)$. 并称它为 $\xi$ 关于 $\eta$ 的条件期望,它是 $\eta$ 的一个函数,因此它是随机变量,即

$$E(\xi|\eta)=\frac{\sum_i x_i p_{x_i,\eta}}{\sum_i p_{x_i,\eta}}$$

(当 $(\xi,\eta)$ 有离散联合分布),其中

$$p_{x_i,y_j}\triangleq P(\xi=x_i,\eta=y_j);$$

$$E(\xi|\eta)=\frac{\int xf(x,\eta)\mathrm{d}x}{f_2(\eta)}\quad(\text{当}(\xi,\eta)\text{ 有密度 } f),$$

此处 $f_2(y)=\int f(x,y)\mathrm{d}x$.

条件期望有两个重要的特征性质:

$(E_1)$　$E(\xi|\eta)$ 是 $\eta$ 的函数;

$(E_2)$　对于任意随机事件 $A$,有 $E(E(\xi|\eta)I_{\eta\in A})=E(\xi I_{\eta\in A})$.

**证明**　$(E_2)$ 若 $(\xi,\eta)$ 离散,则

$$\begin{aligned}E(E(\xi|\eta)I_{\eta\in A})&=E(\varphi(\eta)I_A(\eta))=\sum_j\varphi(y_j)I_A(y_j)\Big/\sum_i p_{x_i,y_j}\\&=\sum_{y_j\in A}\frac{\sum_i x_i p_{x_i,y_j}}{\sum_i p_{x_i,y_j}}\sum_i p_{x_i,y_j}=\sum_i\sum_{y_j\in A}x_i p_{x_i,y_j}\\&=E(\xi I_A(\eta))=E(\xi I_{\eta\in A}).\end{aligned}$$

类似地可以证明:当 $(\xi,\eta)$ 有密度 $f(x,y)$ 时,$(E_2)$ 仍然成立.

$(E_2)$ 有一个等价形式:

$(E_2')$ 对于任意"好的"函数 $g$(这种函数可以通过连续函数列通过"多次"极限过程得到，称为 Borel 函数)均有

$$E(E(\xi|\eta)g(\eta)) = E(\xi g(\eta)).$$

显见 $(E_2)$ 是 $(E_2')$ 的特殊情形，这时 $g(y) = I_A(y)$，而从 $(E_2)$ 到 $(E_2')$ 则可以通过(测度论的)极限过程得到.

再则，如果 $\psi(\eta)$ 满足：对任意 Borel 函数 $g$ 均有 $E(\psi(\eta)g(\eta)) = E(\xi g(\eta))$，则对 $\varphi(\eta)=E(\xi|\eta)$ 必须有

$$P(\psi(\eta) \neq E(\xi|\eta)) = 0.$$

这是因为：若

$$E(\psi(\eta)g(\eta)) = E(\xi g(\eta)) = E(E(\xi|\eta)g(\eta)) = E(\varphi(\eta)g(\eta)),$$

取 $g=\psi-\varphi$，便得 $E(\psi(\eta)-\varphi(\eta))^2=0$. 因此 $\psi(\eta)$ 与 $\varphi(\eta)$ 概率为 1 地相等.

由此可见如果 $\psi(\eta)$ 满足

$$E(\psi(\eta)g(\eta)) = E(\xi g(\eta)) \quad (\forall \text{ Borel 函数 } g).$$

那么 $E(\xi|\eta)$ 与 $\psi(\eta)$ 只差零概率事件. 我们容许 $E(\xi|\eta)$ 差一个零概率事件，即若 $P(\psi(\eta)\neq E(\xi|\eta))=0$，则总认为 $\psi(\eta)=E(\xi|\eta)$ (不计零概率).

这样，我们可以说 $E(\xi|\eta)$ 满足下两个条件的唯一解：

$(E_1'')$ 是 $\eta$ 的函数 $\psi(\eta)$；

$(E_2'')$ $E(\psi(\eta)I_{\eta\in A})=E(\xi I_{\eta\in A})$ ($\forall$ 随机事件 $A$).

当 $(\xi,\eta)$ 既非离散型，又没有联合密度时，初等的定义就没有了. 然而即使在简单的情形，也常会遇见. 例如当 $\xi=h(\eta)$ 且 $\eta$ 非离散的情形. 但是在直观上显然应该有 $E(h(\eta)|\eta)=h(\eta)$. 如果我们对不管什么类型的 $(\xi,\eta)$，均用满足 $(E_1'')$，$(E_2'')$ 的 $\psi(\eta)$ 作为 $E(\xi|\eta)$ 的推广了的统一的定义，那么显然地有 $E(h(\eta)|\eta)=h(\eta)$.

可见对一般的 $(\xi,\eta)$ 应该定义 $E(\xi|\eta)$ 为满足 $(E_1'')$，$(E_2'')$ 的解. 这个解必定存在. 但是要证明这一点就需要用测度论的工具，本书不再涉及.

我们还可以在离散与有密度情形验证以下等式(事实上利用

$(E_1')$ 及 $(E_2')$ 可以证明这些等式对于一般情形也正确)：

(1) $E(E(\xi|\eta))=E(\eta)$；

(2) 若 $\xi$, $\eta$ 独立，则 $E(\xi|\eta)=E(\xi)$；

(3) $E(h(\eta)\xi|\eta)=h(\eta)E(\xi|\eta)$；

(4) $E(\alpha\xi_1+\beta\xi_2|\eta)=\alpha E(\xi_1|\eta)+\beta E(\xi_2|\eta)$.

最后，还可以推广到对多个随机变量的条件期望. 定义

$$E(\xi|\eta_1,\cdots,\eta_n)=\varphi(\eta_1,\cdots,\eta_n),$$

其中

$$\varphi(y_1,\cdots,y_n)=E(\xi|\eta_1=y_1,\cdots,\eta_n=y_n),$$

同样也有

(1) $E(E(\xi|\eta_1,\cdots,\eta_n))=E(\xi)$；

(2) 若 $\xi$ 与 $(\eta_1,\cdots,\eta_n)$ 独立，则 $E(\xi|\eta_1,\cdots,\eta_n)=E(\xi)$；

(3) $E(h(\eta_1,\cdots,\eta_n)\xi|\eta_1,\cdots,\eta_n)$
$$=h(\eta_1,\cdots,\eta_n)E(\xi|\eta_1,\cdots,\eta_n);$$

(4) $E\left(\sum_{i=1}^{l}\alpha_i\xi_i|\eta_1,\cdots,\eta_n\right)=\sum_{i=1}^{l}\alpha_i E(\xi_i|\eta_1,\cdots,\eta_n)$；

(5) 若 $\psi(\eta,a)\triangleq E(h(a,\xi)|\eta)$；则 $E(h(\eta,\xi)|\eta)=\psi(\eta,\eta)$.

特别，若 $\xi$, $\eta$ 独立，则 $E(h(\xi,\eta)|\eta)=g(\eta)$，其中 $g(a)\triangleq Eh(a,\xi)$.

**应用** 设 $\xi$ 不可观测，$\eta$ 可观测，希望用 $\eta$ 的某个函数 $h(\eta)$ 作为 $\xi$ 的预测，使

$$E(\xi-h(\eta))^2=\min_g E(\xi-g(\eta))^2.$$

我们来证明这个最佳预测是 $h(\eta)=E(\xi|\eta)$. 由命题 8.2，这只需验证对任意 $g(\cdot)$ 有

$$E(\xi-E(\xi|\eta)|g(\eta))=0.$$

注意用性质(3)及性质(1)，我们得到

$$\begin{aligned}左&=E(\xi g(\eta))-E(E(\xi|\eta)g(\eta))\\&=E(\xi g(\eta))-E(E(g(\eta)\xi|\eta))\\&=E(\xi g(\eta))-E(g(\eta)\xi)=0.\end{aligned}$$

这就验证了：在均方误差作为标准下，用 $\eta$ 去预测 $\xi$ 的最佳预测就

是条件期望 $E(\xi|\eta)$.

类似地,用 $\eta_1,\cdots,\eta_n$ 去预测 $\xi$,则最佳均方预测为

$$E(\xi|\eta_1,\cdots,\eta_n).$$

**例 1** $(\xi,\eta)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,这时

$$E(\xi|\eta)=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(\eta-\mu_2).$$

**例 2** (AR 模型)

$$\xi_{n+1}=a_1\xi_n+a_2\xi_{n-1}+\cdots+a_p\xi_{n-p+1}+W_{n+1},$$

其中 $\{W_n\}$ 为独立同分布列, $EW_n=0$. 由于这时 $W_{n+1}$ 与 $(\xi_n,\cdots,\xi_{n-p+1})$ 独立,故

$$\begin{aligned}
&E(\xi_{n+1}|\xi_n,\xi_{n-1}\cdots,\xi_{n-p-1})\\
&\quad=E(a_1\xi_n+a_2\xi_{n-1}+\cdots+a_p\xi_{n-p+1}|\xi_n,\xi_{n-1}\cdots,\xi_{n-p-1})+EW_{n+1}\\
&\quad=a_1\xi_n+a_2\xi_{n-1}+\cdots+a_p\xi_{n-p+1}.
\end{aligned}$$

## 习　　题

1. 对 Gauss 系 $\xi$ 定义相关矩阵 $B(s,t)\triangleq E\xi_s\xi_t$,期望函数为 $m(t)=E\xi_t$. 证明

1) $B(t,s)$ 为非负定的,即任意 $n,t_1,\cdots,t_n$,矩阵 $(B(t_i,t_j))_{i,j\leqslant n}$ 为对称非负定(即对于任意实数 $\alpha_1,\cdots,\alpha_n$,恒有

$$\sum_{i,j=1}^{n}B(t_i,t_j)\alpha_i\alpha_j\geqslant 0.$$

对复 Gauss 系 $\zeta$,则 $R(s,t)\triangleq E\zeta_s\bar{\zeta}_t$ 是 Hermit 非负定,即 $(R(t_i,t_j))=(R(t_i,t_j))^*$ (指共轭转置)且对于任意复数 $\alpha_1,\cdots,\alpha_n$,恒有

$$\sum_{i,j=1}^{n}R(t_i,t_j)\alpha_i\bar{\alpha}_j\geqslant 0.$$

2) $\xi$ 的分布(指一切有限维分布)由 $m(t)$ 及 $B(t,s)$ 完全确定(对复 Gauss 系也有相应有结论).

2. 对复 Gauss 系证明

$$E\zeta_t^2 = 0, E(\zeta_t\zeta_s) = E\zeta_t E\zeta_s.$$

3. 复 Gauss 系 $\zeta=\xi+\mathrm{i}\eta$ 有实部与虚部合起来:$\xi\cup\eta$,是实 Gauss 系.

4. 复 Gauss 系 $\zeta$ 相互正交(指所有的元素 $\zeta_t$,$\zeta_s$ 都相互正交)的充要条件是相互独立(指 $\forall n$ 及 $t_1,\cdots,t_n$,二维随机向量族 $\begin{pmatrix}\xi_{t_1}\\ \eta_{t_1}\end{pmatrix}$,$\cdots$,$\begin{pmatrix}\xi_{t_n}\\ \eta_{t_n}\end{pmatrix}$ 恒为相互独立).

5. 试对复 Gauss 系 $\zeta$ 定义 $L(\zeta)$,并证明它仍是复 Gauss 系.

6. 对下列情况举例:

(1) 两个 Gauss 随机变量的和可以不是 Gauss 随机变量.

(2) 两个 Gauss 随机变量不相关但可以不独立.

7. 证明(Stein 引理)$\xi\sim N(0,1)$的充要条件是:对任意可微函数 $f$,只要 $\int|f'(x)|\mathrm{e}^{(-x^2/2)}\mathrm{d}x<\infty$,就有

$$Ef'(\xi) = E(\xi f(\xi)).$$

提示:必要性用

$$\mathrm{e}^{\left(-\frac{x^2}{2}\right)} = \begin{cases}\int_{x}^{\infty} y\mathrm{e}^{(-y^2/2)}\mathrm{d}y, & x\geqslant 0;\\ -\int_{-\infty}^{x} y\mathrm{e}^{(-y^2/2)}\mathrm{d}y, & x<0.\end{cases}$$

充分性:令 $f$ 为 $f'=xf+(I_{(-\infty,x_0]}(x)-\Phi(x_0))$的解,其中

$$\Phi(x) = \int_{-\infty}^{x} \mathrm{e}^{-\frac{y^2}{2}}\mathrm{d}y.$$

8. 若 $\theta_k$ 独立且均遵从$[0,2\pi]$上均匀分布,问

$$\xi_n = \sum_{k=1}^{m} a_k\cos(\lambda_k n + \theta_k)$$

是否宽平稳,是否平稳?

9. 复随机列

$$\xi_n = \sum_{k=1}^{m}\eta_k\mathrm{e}^{\mathrm{i}\lambda_k n}\quad(\lambda_k\ 两两不相等)$$

$(E\eta_k=0, E|\eta_k|^2<\infty, \eta_k=\eta_k'+i\eta_k'')$什么时候宽平稳?

10. 若 $\sum_{k=0}^{\infty}|a_k|<\infty$, $\xi_n=\sum_{k=0}^{\infty}a_k\omega_{n-k}$, $\{\omega_n\}$ 为均值 0 的 i.i.d. 列, $E\omega_n^2=1$, 证明:

(1) $\xi_n$ 为宽平稳.

(2) $R(n)=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{in\lambda}\left|\sum_{k=0}^{\infty}a_k e^{-ik\lambda}\right|^2 d\lambda.$

11. 若 $\xi_n$ 为 MA($q$), 则

$$R(n)=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{in\lambda}|\beta(e^{-i\lambda})|^2 d\lambda.$$

注: 此条件实际上也充分.

12. 若宽平稳列 $\xi_n$ 满足 $E\xi_n=0$, 且

$$R(n)=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{in\lambda}\frac{1}{|\alpha(e^{-i\lambda})|^2}d\lambda,$$

其中 $\alpha(z)$ 的零根均在 $|z|>1$ 中, 则 $\xi_n$ 是 AR.

注: 此条件实际上也必要.

13. $\{\omega_n\}$ 为 i.i.d., $E\omega_n=0$, $E\omega_n^2=1$, $\xi_n=a+bn+\omega_n$, 问 $\eta_n$ 的 $m$-平均:

$$\eta_n \triangleq \frac{1}{2m+1}\sum_{k=-m}^{m}\xi_{n+k}$$

是否宽平稳? 是否平稳?

14. $\{\omega_n\}$ 同 6, 问以下哪些是宽平稳的? 相关函数是什么?

(1) $\xi_n=\omega_n\cos\lambda t+\omega_{n+1}\sin\lambda t$;

(2) $\xi_n=\omega_n\omega_{n+1}$;

(3) $\xi_n=\omega_n\cos(\lambda n)$.

15. 设 $\{\xi_n\}$ 为正态列, $E\xi_n=0(\forall n\geqslant 0)$, 证明

$$E(\xi_i\xi_j\xi_k\xi_l)=\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk},$$

其中 $\sigma_{ij}\triangleq E(\xi_i\xi_j)$. 如果再设 $\{\xi_n\}$ 为 ARMA, 记 $R(n)=E(\xi_k\xi_{k+n})$, 令

$$\hat{R}_n = \hat{R}_{-n} \triangleq \frac{1}{N} \sum_{t=1}^{N-n} \xi_t \xi_{t+n} \quad (n = 0, 1, \cdots, N-1),$$

证明

$$E(\hat{R}_n - R(n))(\hat{R}_m - R(m))$$

$$\frac{1}{N} \sum_{k} (R(k-n)R(k-m) + R(k-n)R(k+m)) + O\left(\frac{1}{N^2}\right).$$

# 第九章　Markov 过程与随机微分方程

## §9.1　平稳 Gauss 过程与 Markov 过程

### 1. 平稳 Gauss 过程

**定义 9.1**　宽平稳过程$\{\xi_t, t\in R\}$称为**宽马氏**的，如果其相关函数 $B(s)(\triangleq E\xi_t\xi_{t+s})$连续，而且对任意 $s\geqslant 0, t\in R$ 有

$$\inf_{n,\, a_i,\, t_i\le t,\, a_i\in R} E|\xi_{s+t}-\sum a_i\xi_{t_i}|^2=\inf_{a\in R}E|\xi_{s+t}-a\xi_t|^2. \tag{9.1}$$

意即：在均方距离(内积$(\xi,\eta)\triangleq E(\xi\eta)$对应的距离)下，用过程在 $t$ 前的资料$\{\xi_u: u\le t\}$的有限线性组合去近似 $\xi_{t+s}$，其最佳的近似只需在 $L(t)\triangleq\{a\xi_t: a\in R^1\}$中找.

对于宽马氏的宽平稳过程，(9.1)右端必然在某个(依赖于 $s$ 的)$a$ 值达到. 我们记此 $a$ 为 $\alpha(s)$. 再则，既然(9.1)成立，那么对任意 $u\geqslant 0$ 及 $\lambda_1,\lambda_2$ 均有

$$E|\xi_{s+t}-(\lambda_1\xi_{t-u}+\lambda_2\xi_t)|^2\geqslant E|\xi_{s+t}-\alpha(s)\xi_t|^2.$$

从而对$\forall s,u\geqslant 0$，由命题 8.2 有

$$(\xi_{s+t}-\alpha(s)\xi_t,\xi_{t-u})=E((\xi_{s+t}-\alpha(s)\xi_t)\xi_{t-u})=0,$$

即

$$B(s+u)=\alpha(s)B(u). \tag{9.2}$$

取 $u=0$，我们得到 $\alpha(s)=B(s)/B(0)$. 于是在(9.2)两端除以 $B(0)$ 后便得到

$$\alpha(s+u)=\alpha(s)\alpha(u)\quad(s,u\geqslant 0).$$

利用 $B(s)$的连续性，即 $\alpha(s)$的连续性，及 $\alpha(s)\leqslant 1=\alpha(0)$，我们有

$$\alpha(s)=e^{-\beta s}\quad(\beta\geqslant 0, s\geqslant 0).$$

这样（记 $\gamma = B(0) \geqslant 0$） $B(s) = \gamma e^{-\beta s}$.

注意，$\gamma=0$ 表示 $E\xi_t^2=0$，即 $\xi_t$ 以概率为 1 地为 0，而 $\beta=0$ 表示 $E\xi_t^2=E(\xi_t\xi_s)$. 于是 $E(\xi_t\xi_s)=(E\xi_t^2\cdot E\xi_s^2)^{1/2}$，这是 Schwarz 不等式的等式情形，因此概率为 1 地 $\xi_s$ 与 $\xi_t$ 成比例，即存在常数 $a_s$，使 $\xi_s=a_s\xi_t$，但是由宽平稳性，$E\xi_t$ 应是常数 $m$，所以 $m=a_s m$，即 $a_s=1$，从而 $\xi_s\equiv\xi_t$. 从而以上 $\gamma=0$ 及 $\beta=0$ 的情形都是退化情形. 对于非退化情形我们便得到

$$B(s)=\gamma e^{-\beta s}\quad(\gamma,\beta>0, s\geqslant 0).$$

再则，当 $s<0$ 时，我们显见有（注意这里是实值一维宽平稳过程）$B(s)=B(-s)$，于是 $B(s)=\gamma e^{\beta s}$ $(s<0)$. 合起来便有

$$B(s)=\gamma e^{-\beta|s|}\quad(\gamma,\beta>0). \tag{9.3}$$

**命题 9.1** 对于具有连续相关函数 $B(t)$ 的实值非退化宽平稳过程 $\{\xi_t, t\in R\}$，它是宽马氏的充要条件为

$$B(s)=\gamma e^{-\beta|s|}\quad(\gamma,\beta>0). \tag{9.3}$$

**证明** 必要性已证. 下面证明充分性，不妨设 $s\geqslant 0$. 这时由(9.2)对于任意 $u\geqslant 0$ 有

$$\begin{aligned}(\xi_{t+s}-e^{-\beta s}\xi_t, \xi_{t-u}) &= B(s+u)-e^{-\beta s}B(u)\\ &=\gamma e^{-\beta(u+s)}-e^{-\beta s}\gamma e^{-\beta u}=0.\end{aligned} \tag{9.4}$$

在 $u=0$ 时，由(9.4)及命题 8.2 得

$$E|\xi_{t+s}-e^{-\beta s}\xi_t|^2=\inf_{a\in R}E|\xi_{t+s}-a\xi_t|^2.$$

再任取 $n$ 并分别令 $t-u=t_1,\cdots,t_n,t$，由(9.4)我们得到

$$\begin{aligned}E|\xi_{t+s}-e^{-\beta s}\xi_t|^2 &= \inf_{a_1,\cdots,a_n,a}E\Big|\xi_{t+s}-\Big(\sum a_i\xi_{t_i}+a\xi_t\Big)\Big|^2\\ &\leqslant \inf_{a_1,\cdots,a_n}E\Big|\xi_{t+s}-\sum a_i\xi_{t_i}\Big|^2.\end{aligned}$$

又由于 $t_1,\cdots,t_n\leqslant t$ 可以任意，$n$ 也可以任意，所以

$$E|\xi_{t+s}-e^{-\beta s}\xi_t|^2=\inf_{n,t_i\leqslant t,1\leqslant i\leqslant n,a_i\in R}E\Big|\xi_{t+s}-\sum a_i\xi_{t_i}\Big|^2.$$

从而(9.1)成立.

**命题 9.2** 宽平稳 Gauss 过程是平稳过程.

**证明** 设$\{\xi_t: t\in R\}$宽平稳，$E\xi_t = m$，$B(t) = E(\xi_{t+s}\xi_s)$，那么由Gauss性$(\xi_{t_1+t},\cdots,\xi_{t_n+t})$的联合特征函数为

$$f(\lambda_1,\cdots,\lambda_n) \triangleq e^{i\sum m\lambda_i - \frac{1}{2}\lambda\sum\lambda^T},$$

其中$\lambda=(\lambda_1,\cdots,\lambda_n)$，$\lambda^T$为它的转置，而$\sum = (R(t_j - t_k))_{j,k\leqslant n}$，$R(t) = B(t) - m^2$. 由于$\sum$只依赖$t_j(j\leqslant n)$而与$t$无关，因此$(\xi_{t_1+t},\cdots,\xi_{t_n+t})$的联合分布函数就与$t$无关. 这说明$\{\xi_t: t\in R\}$是平稳过程.

**命题 9.3** 非退化的宽马氏的平稳Gauss过程$\{\xi_t: t\in R\}$(定义见(9.3))在任意有限个时刻所得的随机向量$(\xi_{t_1},\cdots,\xi_{t_n})$必具有联合分布密度.

**证明** 不妨假定$E\xi_t=0$. 由命题9.1本命题等价于：对任意$n, t_1,\cdots,t_n$，矩阵

$$(B(t_i - t_j))_{i,j\leqslant n} = \gamma(e^{-\beta|t_i - t_j|})_{i,j\leqslant n}$$

为非退化的，即它是可逆矩阵. 为使记号简单我们不失一般地假定$\gamma=\beta=1$，$t_1<\cdots<t_n$. 由(9.1)及(9.2)我们得到

$$(\xi_{t_n} - e^{-(t_n - t_{n-1})}\xi_{t_{n-1}}, \xi_{t_i}) = 0 \quad (i\leqslant n-1).$$

利用$\{\xi_t: t\in R\}$的Gauss性就有：$\xi_{t_n}-e^{-(t_n-t_{n-1})}\xi_{t_{n-1}}$与$(\xi_{t_{n-1}},\cdots,\xi_{t_1})$相互独立. 现在我们用数学归纳法证明$(\xi_{t_1},\cdots,\xi_{t_n})$有联合密度. 当$n=1$时，$\xi_{t_1}\sim N(0,1)$显然有密度，今假定对$n=m-1$时有密度，并记它为$\rho(x_1,\cdots,x_{m-1})$. 往证对于$n=m$结论也成立. 为此我们再记$\eta_m\triangleq\xi_{t_m}-e^{-(t_m-t_{m-1})}\xi_{t_{m-1}}$，其方差为

$$\begin{aligned}\mathrm{Var}\,\eta_m &= E(\xi_{t_m} - e^{-(t_m-t_{m-1})}\xi_{t_{m-1}})^2 \\ &= B(0) - 2e^{-(t_m-t_{m-1})}B(t_m - t_{m-1}) + e^{-2(t_m-t_{m-1})}B(0) \\ &= 1 - e^{-2(t_m-t_{m-1})},\end{aligned}$$

即$\eta_m\sim N(0,1-e^{-2(t_m-t_{m-1})})$. 由于它与$(\xi_{t_{m-1}},\cdots,\xi_{t_1})$独立，故$(\eta_m,\xi_{t_{m-1}},\cdots,\xi_{t_1})$有密度：

$$\rho(x_1,\cdots,x_{m-1})\left[2\pi(1-e^{-2(t_m-t_{m-1})})\right]^{-\frac{1}{2}}e^{-\frac{x_m^2}{2(1-e^{-2(t_m-t_{m-1})})}}.$$

于是$(\xi_{t_m},\cdots,\xi_{t_1})$就有密度

$$\rho(x_1,\cdots,x_m)=\rho(x_1,\cdots,x_{m-1})$$

$$\times\frac{1}{\sqrt{2\pi\left(1-e^{-2(t_m-t_{m-1})}\right)}}e^{-\frac{(x_m-e^{-2(t_m-t_{m-1})}x_{m-1})^2}{2(1-e^{-2(t_m-t_{m-1})})}}. \tag{9.5}$$

**推论** $(\xi_{t_1},\cdots,\xi_{t_n})$的联合分布密度为

$$\rho(x_1,\cdots,x_n)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\prod_{m=2}^{n}\frac{e^{-\frac{(x_m-e^{-2(t_m-t_{m-1})}x_{m-1})^2}{2(1-e^{-2(t_m-t_{m-1})})}}}{\sqrt{2\pi\left(1-e^{-2(t_m-t_{m-1})}\right)}}.$$

**证明** 由命题 9.4 用归纳法即得.

**命题 9.4** 非退化的宽马氏的平稳 Gauss 过程$\{\xi_t:t\in R\}$在任意 $n+1$ 个时刻 $t_1<t_2<\cdots<t_n<t_{n+1}$,$(\xi_{t_1},\cdots,\xi_{t_m})$的联合分布密度记为 $\rho_{t_1,\cdots,t_m}(x_{t_1},\cdots,x_{t_m})$,它具有以下性质:

$$\frac{\rho_{t_1,\cdots,t_{n+1}}(x_1,\cdots,x_{n+1})}{\rho_{t_1,\cdots,t_n}(x_1,\cdots,x_n)}=\frac{\rho_{t_n,t_{n+1}}(x_n,x_{n+1})}{\rho_{t_n}(x_n)}. \tag{9.6}$$

**证明** 由命题 9.3 的证明可见左端为

$$\frac{e^{-\frac{(x_m-e^{-2(t_m-t_{m-1})}x_{m-1})^2}{2(1-e^{-2(t_m-t_{m-1})})}}}{\sqrt{2\pi\left(1-e^{-2(t_m-t_{m-1})}\right)}}.$$

这个表达式与 $t_1,\cdots,t_{n-1},x_1,\cdots,x_{n-1}$无关,所以等式成立.

命题 9.4 的含义为:在给定 $\xi_{t_n}=x_n,\cdots,\xi_{t_1}=x_1$ 的条件下,$\xi_{t_{n+1}}$的条件分布密度(即(9.6)中的左端)与 $\xi_{t_{n-1}}=x_{n-1},\cdots,\xi_{t_1}=x_1$ 无关,它就是 $\xi_{t_{n+1}}$在给定的 $\xi_{t_n}=x_n$ 的条件下的条件分布密度,换句话说,在给定条件 $\xi_{t_n}=x_n$ 的条件下,$\xi_{t_{n+1}}$与$(\xi_{t_{n-1}},\cdots,\xi_{t_1})$独立.

我们把(9.6)两端的条件分布密度分别记为

$$p_{\xi_{t_{n+1}}}(x_{n+1}|\xi_{t_n}=x_n,\cdots,\xi_{t_1}=x_1)\quad 及\quad p_{\xi_{t_{n+1}}}(x_{n+1}|\xi_{t_n}=x_n),$$

那么(9.4)即

$$p_{\xi_{t_{n+1}}}(x_{n+1}|\xi_{t_n}=x_n,\cdots,\xi_{t_1}=x_1)=p_{\xi_{t_{n+1}}}(x_{n+1}|\xi_{t_n}=x_n).$$

进一步对 $x_{n+1}$在区域(或集合)$A$ 上作积分,于是我们有

$$P(\xi_{t_{n+1}} \in A \mid \xi_{t_n} = x_n, \cdots, \xi_{t_1} = x_1) = P(\xi_{t_{n+1}} \in A \mid \xi_{t_n} = x_n).$$

$\{\xi_t; t \in R\}$的这个性质称为 **Markov 性**，它表明已知现在($t=t_n$)时刻过程的取值 $\xi_{t_n}=x_n$ 的条件下，过去的取值 $\xi_{t_{n-1}}=x_{n-1}$ 等并不影响将来($t=s>t_n$)的 $\xi_{t_{n+1}}$ 取值的统计规律. 这里的 Markov 性，正是第 6 章中的离散时间的连续状态 Markov 性到连续时间的推广.

利用(9.6)及归纳法我们可以把命题 9.4 的推论改写为：$\xi_{t_1}$，$\cdots$，$\xi_{t_n}$的联合密度

$$p_{\xi_{t_1},\cdots,\xi_{t_n}}(x_1,\cdots,x_n)$$

$$= p_{\xi_{t_1}}(x_1) p_{\xi_{t_2}}(x_2 \mid \xi_{t_1} = x_1) \cdots p_{\xi_{t_n}}(x_n \mid \xi_{t_{n-1}} = x_{n-1}). \qquad (9.5)'$$

它即是“初始”密度 $p_{\xi_{t_1}}(x_1)$ 与一系列条件密度的乘积.

以下我们只考虑 $t \geqslant 0$. 即考虑不退化的宽马氏平稳 Gauss 过程$\{\xi_t; t \geqslant 0\}$. 并设 $E\xi_t = 0, B(t) = \gamma e^{-\beta|t|}$ $(\gamma, \beta > 0)$. 改记($s \geqslant 0, t > 0$)：

$$p(t,x,y) \triangleq p_{\xi_{t+s}}(y \mid \xi_s = x) = \frac{e^{-\frac{(y - x e^{-\beta t})^2}{2\gamma(1 - e^{-2\beta t})}}}{\sqrt{2\pi\gamma(1 - e^{-2\beta t})}}$$

(注意上述右端与 $s$ 无关).

**命题 9.5** $\{p(t,x,y)\}$满足 Kolmogorov-Chapman 方程：

$$p(t+s,x,y) = \int p(s,x,z) p(t,z,y) \mathrm{d}z. \qquad (9.7)$$

**证明** 为对称起见，令 $t_1+s=t_2, t_2+t=t_3, x=x_1, z=x_2, y=x_3$，于是上式

$$右端 = \int p_{\xi_{t_2}}(x_2 \mid \xi_{t_1} = x_1) p_{\xi_{t_3}}(x_3 \mid \xi_{t_2} = x_2) \mathrm{d}x_2$$

$$= \int \frac{p_{\xi_{t_1},\xi_{t_2}}(x_1,x_2)}{p_{\xi_{t_1}}(x_1)} \frac{p_{\xi_1,\xi_2,\xi_3}(x_1,x_2,x_3)}{p_{\xi_1,\xi_2}(x_1,x_2)} \mathrm{d}x_2$$

$$= \frac{p_{\xi_{t_1},\xi_{t_3}}(x_1,x_3)}{p_{\xi_{t_1}}(x_1)} = p_{\xi_{t_3}}(x_3 \mid \xi_{t_1} = x_1) = 左端.$$

## 2. 实值时齐 Markov 过程

**定义 9.2** 设随机过程$\{\xi_t;t>0\}$的任意有限个时刻所对应的随机向量$(\xi_{t_1},\cdots,\xi_{t_n})$都有正的联合分布密度函数，而条件分布密度满足对任意的$t_1<\cdots<t_n$及$t\geqslant 0$都有

$$
\begin{aligned}
& p_{\xi_{t_n+t}}(y|\xi_{t_1}=x_1,\cdots,\xi_{t_{n-1}}=x_{n-1},\xi_{t_n}=x) \\
&= p_{\xi_{t_n+t}}(y|\xi_{t_n}=x)(\triangleq p(t,x,y)),
\end{aligned}
\tag{9.8}
$$

则称$\{\xi_t;t\geqslant 0\}$为实值**时齐的 Markov 过程**，又称$p(t,x,y)$为这个时齐的 Markov 过程的**转移密度函数**.

**注 1**：(9.7)是§5.2条件$(P_1)$在状态空间为实数集时的对应关系(差别只在于级数换成了积分).

显见，对于实值时齐的 Markov 过程$\{\xi_t;t\geqslant 0\}$恒有：对任意的$s>s_m>\cdots>s_1,t>0,u>0$,

$$
\begin{aligned}
& P(\xi_{s+u}\in A|\xi_s=x,\xi_{s_m}=x_m,\cdots,\xi_{s_1}=x_1) \\
&= P(\xi_{s+u}\in A|\xi_s=x)=P(\xi_{t+u}\in A|\xi_t=x).
\end{aligned}
\tag{9.8'}
$$

**注 2**：定义 9.2 中关于存在正的有限维联合分布密度的要求不是必须的. 实际上，只需要用(9.8)′代替(9.8)作为定义就足够了. 但是这就需要用测度论的知识，在本书中从略.

**注 3**：定义 9.2 中的时齐 Markov 过程虽然只在$t>0$时给出了定义，但是我们可以自然地利用时齐性补充定义$\xi_0=x$及$p_{\xi_t}(y|\xi_0=x)\triangleq p_{\xi_{t+s}}(y|\xi_s=x)$ $(t,s>0$，上式右端不依赖于具体的$s)$. 从而可以讨论$\xi_0$取离散值或有密度的情形(如果利用较多一点的数学知识——Stieltjes 积分，则$\xi_0$可取任意的随机变量). 也就是说时齐 Markov 过程$\{\xi_t;t>0\}$总可以扩展为时齐的 Markov 过程$\{\xi_t;t\geqslant 0\}$(加入$t=0$的$\xi_0$).

**例 9.1** 非退化的宽马氏平稳 Gauss 过程是时齐的 Markov 过程. 退化的宽马氏平稳 Gauss 过程(即$\gamma=0$或$\beta=0$的情形)$\{\xi_t;t\geqslant 0\}$，则一定概率为 1 地有$\xi_t=\xi_0$，其中$\xi_0$是一个 Gauss 随机变量(正态随机变量或常数). 这时$(\xi_{t_1},\cdots,\xi_{t_n})$并不存在联合分布密度，我们的定义 9.2 并不包含这种情形. 但是注 2 中所述的定义

就自然地包含了这种情形.这样我们就能说:对于平稳 Gauss 过程来说宽马氏性与马氏性(即 Markov 性)是一样的.

**例 9.2** §2.2 中定义的 Brown 运动是时齐的实值 Markov 过程(作为习题).

**定义 9.3** 转移函数为

$$p(t,x,y) = p_{\xi_{s+t}}(y|\xi_s = x) = \frac{\exp\left\{-\frac{(y - xe^{-\beta t})^2}{2\gamma(1 - e^{-2\beta t})}\right\}}{(2\pi\gamma(1 - e^{-2\beta t}))^{\frac{1}{2}}}.$$

时齐 Markov 过程称为(参数$(\beta,\gamma)$的)**OU 过程**(Ornstein-Uhlenbeck 过程).

对此我们有

$$\lim_{t\to\infty} p(t,x,y) = \frac{e^{-\frac{y^2}{2\gamma}}}{(2\pi\gamma)^{\frac{1}{2}}}(\triangleq N(0,\gamma)).$$

注意,非退化的宽马氏平稳 Gauss 过程(设 $E\xi_t = 0$)是 OU 过程,而且对任意的 $t$,$\xi_t$ 的分布恒为 $N(0,B(0)) = N(0,\gamma)$. 于是由

$$p_{\xi_t}(y) = \int p_{\xi_0,\xi_t}(x,y)\mathrm{d}x = \int p_{\xi_t}(y|\xi_0 = x)p_{\xi_0}(x)\mathrm{d}x.$$

立得
$$\frac{e^{-\frac{y^2}{2\gamma}}}{\sqrt{2\pi\gamma}} = \int p(t,x,y)\frac{e^{-\frac{x^2}{2\gamma}}}{\sqrt{2\pi\gamma}}\mathrm{d}x.$$

记 $\varphi(x) = \frac{e^{-\frac{x^2}{2\gamma}}}{\sqrt{2\pi\gamma}}$,则上式变为

$$\varphi(y) = \int p(t,x,y)\varphi(x)\mathrm{d}x. \tag{9.9}$$

**定义 9.4** 若一个概率分布密度 $\varphi(x)$ 关于时齐 Markov 过程 $\{\xi_t:t\geqslant 0\}$ 的转移函数 $p(t,x,y)$ 满足(9.9),则称 $\varphi(\cdot)$ 为 $p(t,x,y)$(或 $\{\xi_t:t\geqslant 0\}$)的**不变概率分布密度函数**.

**例 9.3** $N(0,\gamma)$ 是 OU 过程的不变概率分布密度函数. 如果在 OU 过程中置 $\xi_0 \sim N(0,\gamma)$,则我们就得到非退化的马氏平稳 Gauss 过程,可见 OU 过程与平稳 Gauss. Markov 过程的差异只

在于前者在 $t=0$ 时 $\xi_0$ 可以不必局限于遵从 $N(0,Y)$.

本章以下的内容是向读者粗略地介绍一种最常见的连续状态,连续时间的马氏过程——扩散过程及其主要研究工具——随机微积分.由于这部分内容涉及较深的数学内容,要真正表达清楚,测度论与概率基础是必不可少的.这就大大超出了本书大部分读者现有的数学基础.因此,我们在本章中不求一般地给出定义与定理,而是从一些有典型意义的特例来道出随机微积分及扩散过程的概念与主要思想.

## § 9.2 Brown 运动及其微积分

在研究函数的时候,函数的微积分是其精髓.对连续参数的随机过程、随机函数的微积分也有同样的意义.函数的微分与积分是一对互逆的运算.对初等函数,我们常常是由研究其导数入手,进而得到其积分及微分;但是对复杂的一般函数,积分却是比微分较容易搞清的,也易于作近似计算,因此,人们也常以积分作为微积分的核心.下面我们考虑对 Brown 运动的随机积分.

### 1. 随机积分

设 $B\triangleq\{B_t(\omega);t\geqslant 0\}$ 是一个一维 Brown 运动,又设有只依赖 $B$ 的过去的值的随机函数 $\Phi_t\triangleq\{\Phi(t,\omega);t\in(a,b],\omega\in\Omega\}$(即 $\Phi_t$ 只依赖 $\{B_s(\omega);s\leqslant t\}$),又对任意固定的 $\omega$,$\Phi(\cdot,\omega)$ 都是在 $(a,b]$ 上连续的实函数,而且

$$\int_a^b E|\varphi(u,\omega)|^2\mathrm{d}\omega<+\infty.$$

类似于函数微积分中分割求和取极限的逼近方法,我们这里也考虑一列分割:

$$a=t_0^{(n)}<t_1^{(n)}<\cdots<t_{l_n}^{(n)}=b\quad(n=1,2,\cdots);$$

$$\Delta_n=\max_k(t_{k+1}^{(n)}-t_k^{(n)})\to 0\quad(n\to+\infty).$$

取和

$$I_n(\omega) \triangleq \sum_k \Phi(t_k^{(n)},\omega)(B_{t_{k+1}^{(n)}}(\omega) - B_{t_k^{(n)}}(\omega)). \quad (9.10)$$

如果$\{I_n(\omega)\}$对每一个$\omega$都存在极限，那么很自然地，这个极限就该是$\Phi$对$B$的积分，但是，遗憾的是，一般地我们不能保证$\{I_n(\omega)\}$对每个$\omega$都有极限，甚至可能对大部分$\omega$，$I_n(\omega)$都没有实数意义下的极限，所以"极限"的含义就必须改一下. 让我们回忆初等概率论中的概念：设有一个随机变量$\eta(\omega)$使得：对$\forall\ \epsilon>0$，都有

$$\lim_{n\to\infty} P(\omega; |I_n(\omega) - \eta(\omega)| > \epsilon) = 0. \quad (9.11)$$

我们就称$\eta(\omega)$是$I_n(\omega)$依概率意义下的极限. 它的意思是当$n$充分大，$I_n$与$\eta$不充分接近的可能可任意小，记为$I_n(\omega)\xrightarrow[(p)]{}\eta(\omega)$. 可以证明如果$\Phi(t,\omega)$只依赖$\{B_s(\omega); s\leqslant t\}$，那么在依概率意义下，(9.10)定义的$\{I_n\}$有极限，将此极限定义为$\Phi$对$B$在$(a,b]$上的**Ito 随机积分**，并记为

$$\int_a^b \Phi_t \mathrm{d}B_t \triangleq \int_a^b \Phi(t,\cdot)\mathrm{d}B_t \triangleq (p) \lim_{\Delta_n\to\infty} I_n(\omega).$$

一般地，随机积分是不能对每一个固定$\omega$分别定义的. 还要指出一点：在和数$I_n(\omega)$中，每个小区间$(t_k^{(n)},t_{k+1}^{(n)}]$上，我们限定$\Phi(t_k^{(n)},\omega)$取作为$\Phi$在其上的近似，而不能像在普通的函数积分中任取$\Phi$在区间上的一点的值为近似，其原因是：这里相应于在普通积分中$\Delta t_k^{(n)} \triangleq t_{k+1}^{(n)} - t_k^{(n)}$的是$\Delta B_k^{(n)} \triangleq B_{t_{k+1}^{(n)}}(\omega) - B_{t_k^{(n)}}(\omega)$. 后者是取值复杂的随机变量族. 虽然

$$\lim_{n\to\infty} \Delta B_k^{(n)} = 0,$$

但是对于不同的$\omega$，它们趋于0的速度很不一致，而平均地说$\Delta B_k^{(n)} \sim (\Delta t_k^{(n)})^{\frac{1}{2}}$，即

$$E|\Delta B_k^{(n)}|^2 = E|B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}}|^2$$

$$= \int \frac{e^{-\frac{y^2}{2(t_{k+1}^{(n)} - t_k^{(n)})}}}{\sqrt{2\pi(t_{k+1}^{(n)} - t_k^{(n)})}} y^2 \mathrm{d}y$$

$$= t_{k+1}^{(n)} - t_k^{(n)} = \Delta t_k^{(n)}.$$

也就是说，平均地 $\Delta B_k^{(n)}$ 趋于0的速度大大慢于 $\Delta t_k^{(n)}(n \to +\infty)$. 从而在区间上 $(t_k^{(n)}, t_{k+1}^{(n)}]$ 取不同的点的 $\Phi$ 来作近似和，它们之间的差别是不可忽略的(下面我们还将在例子中再见到这一事实). 从而我们必须限定 $\Phi$ 在每个分割小区间上的取值方式.

**例 9.4** $\int_s^t B\mathrm{d}B = \frac{1}{2}(B_t^2 - B_s^2) - \frac{1}{2}(t-s).$

这个结果与微积分的换元公式是不同的，这里多了一项 $-\frac{1}{2}(t-s)$. 下面我们来推导这一结果.

$$\begin{aligned}
I_n &= \sum_k B_{t_k^{(n)}}(B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}}) \\
&= \frac{1}{2}\sum_k [2B_{t_{k+1}^{(n)}}B_{t_k^{(n)}} - 2B_{t_k^{(n)}}^2] \\
&= \frac{1}{2}\sum_k [(B_{t_{k+1}^{(n)}}^2 - B_{t_k^{(n)}}^2) - (B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}})^2] \\
&= \frac{1}{2}(B_t^2 - B_s^2) - \frac{1}{2}\sum_k (B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}})^2 \\
&\triangleq \frac{1}{2}(B_t^2 - B_s^2) - I\!I_n.
\end{aligned}$$

而这里后一项 $I\!I_n$ 的数学期望是

$$\begin{aligned}
E\, I\!I_n &= E\,\frac{1}{2}\sum_k (B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}})^2 = \frac{1}{2}\sum_k E(B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}})^2 \\
&= \frac{1}{2}\sum_k (t_{k+1}^{(n)} - t_k^{(n)}) = \frac{1}{2}(t-s),
\end{aligned}$$

而且它的方差是

$$\begin{aligned}
E\left(I\!I_n - \frac{1}{2}(t-s)\right)^2 &= E\left(\frac{1}{2}\sum_k (B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}})^2 - \frac{1}{2}(t-s)\right)^2 \\
&= \frac{1}{4}\sum_k E[(B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}})^2 - (t_{k+1}^{(n)} - t_k^{(n)})]^2 \\
&= \frac{1}{4}\sum_k (E(B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}})^4 - (t_{k+1}^{(n)} - t_k^{(n)})^2)
\end{aligned}$$

$$= \frac{1}{4}\sum_k \left(3(t_{k+1}^{(n)}-t_k^{(n)})^2-(t_{k+1}^{(n)}-t_k^{(n)})^2\right)$$

$$= \frac{2}{4}\sum_k (t_{k+1}^{(n)}-t_k^{(n)})^2$$

$$\leqslant \frac{1}{2}\Delta_n \sum_k (t_{k+1}^{(n)}-t_k^{(n)})$$

$$= \frac{1}{2}\Delta_n (t-s) \to 0 \quad (n\to+\infty).$$

当 $n\to+\infty$ 时，由 Chebyshev 不等式

$$P\left\{\omega: \left|I_n(\omega)-\frac{1}{2}(t-s)\right|^2 \geqslant \varepsilon\right\} \leqslant \frac{1}{\varepsilon^2}E\left(I_n-\frac{1}{2}(t-s)\right)^2 \to 0.$$

可见

$$P\left\{\omega: \left|I_n-\frac{1}{2}\left[(B_t^2(\omega)-B_s^2(\omega))-(t-s)\right]\right|\geqslant \varepsilon\right\}\to 0.$$

由 Ito 随机积分的定义，此即

$$\int_s^t B_u \mathrm{d}B_u = \frac{1}{2}(B_t^2-B_s^2)-\frac{1}{2}(t-s).$$

下面我们再来看在 $(t_k^{(n)}, t_{k+1}^{(n)}]$ 上取 $\Phi$ 的不同近似点是否会影响积分近似和的极限值. 考查

$$\tilde{I}_n \triangleq \sum_k B_{t_{k+1}^{(n)}}(B_{t_{k+1}^{(n)}}-B_{t_k^{(n)}})$$

$$= \frac{1}{2}\sum_k \left(2B_{t_{k+1}^{(n)}}^2 - 2B_{t_{k+1}^{(n)}}B_{t_k^{(n)}}\right)$$

$$= \frac{1}{2}\sum_k \left[(B_{t_{k+1}^{(n)}}^2 - B_{t_k^{(n)}}^2)+(B_{t_{k+1}^{(n)}}-B_{t_k^{(n)}})^2\right]$$

$$= \frac{1}{2}(B_t^2-B_s^2)+\frac{1}{2}\sum_k (B_{t_{k+1}^{(n)}}-B_{t_k^{(n)}})^2.$$

与前面同理可得

$$P\left\{\omega: \left|\tilde{I}_n-\left[\frac{1}{2}(B_t^2-B_s^2)+\frac{1}{2}(t-s)\right]\right|\geqslant\varepsilon\right\}\to 0 \quad (n\to\infty).$$

可见 $\tilde{I}_n$ 的极限与 $I_n$ 的极限差了 $(t-s)$ 这样一项. 这里让我们再次

强调在随机积分近似和数中被积函数 $\Phi$ 在每个分割小区间上的近似必须限定！Ito 随机积分要求在积分和式中限定取小区间左端的 $\Phi$ 值. 若在每个分割小区间 $(t_k^{(n)}, t_{k+1}^{(n)}]$ 上以 $\frac{1}{2}(\Phi(t_k^{(n)},\omega)+\Phi(t_{k+1}^{(n)},\omega))$ 来近似 $\Phi$, 则有和

$$\hat{I}_n=\sum_k \frac{1}{2}(\Phi(t_k^{(n)},\omega)+\Phi(t_{k+1}^{(n)},\omega))(B_{t_{k+1}^{(n)}}-B_{t_k^{(n)}}). \quad (9.12)$$

而 $\hat{I}_n$ 在依概率意义下的极限就是另一种积分，称为 **Stratonowich 积分**，记为

$$\int_s^t \Phi \circ \mathrm{d}B \triangleq \int_s^t \Phi_u \circ \mathrm{d}B \triangleq (p)\lim_{n\to\infty}\hat{I}_n. \quad (9.13)$$

由前面的讨论，我们容易得到

$$\begin{aligned}\int_s^t B\circ \mathrm{d}B &= \left[\frac{1}{4}(B_t^2-B_s^2)-\frac{1}{4}(t-s)\right] \\ &\quad +\frac{1}{4}[B_t^2-B_s^2+(t-s)] \\ &=\frac{1}{2}(B_t^2-B_s^2)=\frac{1}{2}B_t^2\Big|_s^t.\end{aligned}$$

这个结果与微积分中的结果一致. 这正是 Stratanowich 积分的优点. 但是 Ito 随机积分在计算中有其他更多的优点(如可以利用所谓“鞅论”的工具)，所以在很多情况下，人们更喜欢采用它. 与普通函数的积分类似，随机积分也有下列性质：

Ⅰ. 线性：

$$\int_a^b(\Phi+\Psi)\mathrm{d}B=\int_a^b\Phi\mathrm{d}B+\int_a^b\Psi\mathrm{d}B. \quad (9.14)$$

对实数 $\eta$

$$\int_a^b\eta\Phi_u\mathrm{d}B_u=\eta\int_a^b\Phi_u\mathrm{d}B_u \quad (9.15)$$

(对于 $\eta$ 只依赖于 $\{B_u; u\leqslant a\}$ 的某些随机变量也成立).

Ⅱ. 可加性：

$$\int_a^b\Phi\mathrm{d}B+\int_b^c\Phi\mathrm{d}B=\int_a^c\Phi\mathrm{d}B \quad (9.16)$$

（其实对某些随机时间 $b=\tau(\omega)$——所谓停时，即它对 $\forall t\geqslant 0$，使事件 $\{\omega:\tau(\omega)\leqslant t\}$ 只依赖于 $\{B_u;u\leqslant t\}$，上式仍成立）.

此外作为"随机"积分，它还有性质：

$$\text{Ⅲ}. \quad E\int_a^b \Phi \mathrm{d}B = 0. \tag{9.17}$$

$$\text{Ⅳ}. \quad E\left\{\int_a^b \Phi_u \mathrm{d}B_u \int_a^b \Psi_u \mathrm{d}B_u\right\} = \int_a^b E(\Phi_u \Psi_u)\mathrm{d}u, \tag{9.18}$$

$$E\left|\int_a^b \Phi_u \mathrm{d}B_u\right|^2 = \int_a^b E|\Phi|^2 \mathrm{d}u. \tag{9.19}$$

这些性质的严格证明涉及测度论的许多知识，我们这里只能略去了. 但是为了使读者不致于觉得它们是"天外飞来"的，我们作一些不严格的说明. 由于和对函数是有线性的，因而它们的极限也应有线性，从而可以想象 Ⅰ 会成立. 类似地 Ⅱ 也不难想象应该成立. 至于(9.17)～(9.19)成立的原因就更深一些了. 这里让我们看一个特殊情况：$\Phi,\Psi$ 都是时间的非随机函数 $\Phi_u,\Psi_u$ 的情形，此时我们有

$$EI_n = E\left\{\sum_k \Phi_{t_k^{(n)}}(B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}})\right\}$$

$$= \sum_k \Phi_{t_k^{(n)}} E(B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}}) = 0.$$

从而我们知道，在允许极限交换的情形下

$$E\int_a^b \Phi \mathrm{d}B = E((p)\lim_{n\to\infty} I_n) = 0.$$

另一方面，

$$E\left\{\sum_k \Phi_{t_k^{(n)}}(B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}})\right\}\left\{\sum_l \Psi_{t_l^{(n)}}(B_{t_{l+1}^{(n)}} - B_{t_l^{(n)}})\right\}$$

$$= \sum_{k,l} \Phi_{t_k^{(n)}} \Psi_{t_l^{(n)}} E(B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}})(B_{t_{l+1}^{(n)}} - B_{t_l^{(n)}}).$$

由于 Brown 运动在不交区间上的独立性，上式应等于

$$\sum_k \Phi_{t_k^{(n)}} \Psi_{t_k^{(n)}} E[(B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}})^2]$$

$$+ \sum_{k\neq l} \Phi_{t_k^{(n)}} \Psi_{t_l^{(n)}} E(B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}}) E(B_{t_{l+1}^{(n)}} - B_{t_l^{(n)}})$$

$$= \sum_k \Phi_{t_k^{(n)}} \Psi_{t_k^{(n)}} (t_{k+1}^{(n)} - t_k^{(n)}) \to \int_a^b E(\Phi_u \Psi_u) \mathrm{d}u.$$

## 2. Ito 公式——随机积分的换元公式与复合函数的随机微分公式

在普通函数微积分中，复合函数的微分公式与积分的换元公式是一对等价的最基本的公式. 在随机微积分中，我们也有这样一个对应的公式，从而就使随机积分的概念可以方便地应用到许多问题中去.

若 $\xi_t$ 是由以下等式给定

$$\xi_t = \xi_0 + \int_0^t \sigma(u,\omega) \mathrm{d}B_u(\omega) + \int_0^t b(u,\omega) \mathrm{d}u \qquad (9.20)$$

（也记成 $\mathrm{d}\xi_t = \sigma_u \mathrm{d}B_u + b_u \mathrm{d}u$），其中 $\sigma_t$ 与 $b_t$ 都只依赖 Brown 运动的过去值 $\{B_s; s \leqslant t\}$，而且对任意固定的 $\omega$，都是 $u$ 的连续函数，同时

$$\int_0^T |\sigma(u,\omega)|^2 \mathrm{d}u < +\infty, \quad \int_0^T |b(u,\omega)| \mathrm{d}u < +\infty,$$

其中 $\int_0^T b(u,\omega)\mathrm{d}u$ 理解为固定 $\omega$ 对 $u$ 的普通积分. 又若有二元实函数 $f(t,x)$ 对 $x$ 二阶光滑且对 $t$ 一阶光滑，令

$$\eta_t \triangleq \eta(t,\omega) \triangleq f(t,\xi_t) \qquad (9.21)$$

（$\eta_t$ 是一个复合随机函数）. 现在问：是否有 $\Psi$ 与 $g$ 使得对 $t \in [0, T]$ 都有

$$\eta(t,\omega) = \eta(0,\omega) + \int_0^t \Psi_u \mathrm{d}B_u + \int_0^t g_u \mathrm{d}u,$$

是否能用 $\sigma, b, f$ 来表示 $\Psi$ 与 $g$？

Ito 公式完满地回答了这一问题.

**定理 9.1（Ito 公式）** 对 $x$ 二阶光滑的二元实函数 $f(\cdot,\cdot)$，对(9.11)给定的 $\xi = \{\xi_t; t \in [0,T]\}$ 有

$$\eta_t \triangleq \eta(t,\xi_t) = \eta_0 + \int_0^t \frac{\partial f}{\partial x}(u,\xi_u) \mathrm{d}B_u + \int_0^t \left\{ \frac{\partial f}{\partial t} + b \frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial x^2} \right\} (u,\xi_u) \mathrm{d}u; \quad (9.22)$$

或以微分形式写为：

$$d\eta_t = \frac{\partial f}{\partial x}(t,\xi_t)\sigma dB_t + \left[\frac{\partial f}{\partial t} + b(t,\omega)\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2(t,\omega)\frac{\partial^2 f}{\partial x^2}\right](t,\xi_t)dt.$$

即

$$df(t,\xi_t) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}d\xi_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(d\xi_t)^2,$$

其中 $(d\xi_t)^2 = \sigma^2 dt$.

为了使读者能直观地理解 *Ito* 公式，我们作一个不严格的“推导”.

考察

$$\begin{aligned}\eta_{t+\Delta t} - \eta_t &= f(t+\Delta t,\xi_{t+\Delta t}) - f(t,\xi_t)\\ &= \left[\frac{\partial f}{\partial x}(t,\xi_t)(\xi_{t+\Delta t}-\xi_t) + \frac{\partial f}{\partial t}(t,\xi_t)\Delta t\right]\\ &\quad + \frac{1}{2}\left[\frac{\partial^2 f}{\partial x^2}(t,\xi_t)(\xi_{t+\Delta t}-\xi_t)^2\right.\\ &\quad + 2\frac{\partial^2 f}{\partial x\partial t}(t,\xi_t)(\xi_{t+\Delta t}-\xi_t)\Delta t\\ &\quad \left.+ \frac{\partial^2 f}{\partial t^2}(\Delta t)^2\right] + o(\Delta t).\end{aligned}$$

Taylor 展开，在第二个括号中除第一项外均为 $o(\Delta t)$，但是

$$\begin{aligned}(\xi_{t+\Delta t}-\xi_t)^2 &\doteq (\sigma_t(B_{t+\Delta t}-B_t) + b_t\Delta t)^2\\ &= \sigma_t^2(B_{t+\Delta t}-B_t)^2 + b_t^2(\Delta t)^2 + 2\sigma_t(B_{t+\Delta t}-B_t)\Delta t\\ &\doteq \sigma_t^2\Delta t + o(\Delta t).\end{aligned}$$

于是

$$\begin{aligned}\eta_{t+\Delta t} - \eta_t = &\frac{\partial f}{\partial x}(t,\xi_t)(\sigma(t,\omega)\Delta B_t + b(t,\omega)\Delta t)\\ &+ \frac{\partial f}{\partial t}(t,\xi_t)\Delta t + \frac{1}{2}\left(\frac{\partial^2 f}{\partial x^2}(t,\xi_t)\sigma^2(t,\omega)\Delta t + o(\Delta t),\right.\end{aligned}$$

即有

$$d\eta_t = \frac{\partial f}{\partial x}(t,\xi_t)\sigma(t,\omega)dB_t + \left[\frac{\partial f}{\partial t}(t,\xi_t) + \frac{\partial f}{\partial x}(t,\xi_t)b(t,\omega)\right.$$

$$+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t,\xi_t)\sigma^2(t,\omega)\Big)\mathrm{d}t.$$

**例 9.5** 设 $\sigma, b$ 为常数，$\eta_t = \mathrm{e}^{-bt}\left[\sigma\int_0^t \mathrm{e}^{bs}\mathrm{d}B_s + \xi_0\right] \triangleq \mathrm{e}^{-bt}\xi_t$，则

$$\eta_t = \eta_0 + \int_0^t - b\mathrm{e}^{-bs}\xi_s\mathrm{d}s + \int_0^t \mathrm{e}^{-bs}\sigma\mathrm{e}^{bs}\mathrm{d}B_s$$

$$= \eta_0 + \int_0^t - b\eta_s\mathrm{d}s + \int_0^t \sigma\mathrm{d}B_s,$$

即

$$\mathrm{d}\eta_t = - b\eta_t\mathrm{d}t + \sigma\mathrm{d}B_t. \tag{9.23}$$

**注**：这里$\{\eta_t\}$是一个 Gauss 马氏过程，就是 OU 过程.(9.23)实际上可看作一个以$\{\eta_t\}$为未知过程的随机微分方程(参见下一段)，它最早出现于理论物理研究中，在那里称之为 Langevin 方程. 事实上 $\eta_t$ 是参数为 $\beta = b, \gamma = \frac{\sigma^2}{2b}$ 的 OU 过程.

**例 9.6** 设

$$Z_t = \exp\left(\int_0^t \Phi\mathrm{d}B - \frac{1}{2}\int_0^t \Phi_u^2\mathrm{d}u\right) \xlongequal{\text{记成}} \exp\xi_t, \tag{9.24}$$

则由 Ito 公式推出

$$\mathrm{d}Z_t = \mathrm{d}\mathrm{e}^{\xi_t} = \mathrm{e}^{\xi_t}\mathrm{d}\xi_t + \frac{1}{2}\mathrm{e}^{\xi_t}(\mathrm{d}\xi_t)^2,$$

即

$$Z_t = Z_0 + \int_0^t (\exp\xi_s)\Phi_s\mathrm{d}B_s + \int_0^t (\exp\xi_s)\left(-\frac{1}{2}\Phi_s^2\right)\mathrm{d}s$$

$$+ \int_0^t \frac{1}{2}(\exp\xi_s)\Phi_s^2\mathrm{d}s = Z_0 + \int_0^t Z_s\Phi_s\mathrm{d}B_s.$$

由此得到

$$\begin{cases}\mathrm{d}Z_t = Z_t\Phi_t\mathrm{d}B_t,\\ Z_0 = 1.\end{cases} \tag{9.25}$$

**注**：(9.25)也是一个随机微分方程，事实上(9.24)给出的 $Z$ 是它的唯一解.

**例 9.7**(股票的 **Black-Scholes 模型**)

设某股票在 $t$ 的价格为随机变量 $\xi_t$,且存在常数 $b,\sigma(>0)$,使

$$d\xi_t = \xi_t(b\,dt + \sigma dB_t),$$

其中 $b$ 称为预期收益率,$\sigma$ 称为波动率. 用 Ito 公式可验证 $\xi_t = \xi_0\exp\left(\left(b - \frac{\sigma^2}{2}\right) + \sigma B_t\right)$(称为几何 Brown 运动).

Ito 公式成立的范围可以更广得多,但这些成立条件的描述需要测度论的语言. 我们这里只能就较特殊的情况,把 Ito 公式的概要向读者介绍,以期读者能领略随机微积分的韵味. 要真正完全地,严格地懂得与掌握随机微积分,读者必须掌握基本的测度论知识与方法. 具备此基础的读者可参考[G]或[RY]等著作,进一步学习随机微积分.

### 3. 随机微分方程

在 2 中的三个例子已经给出了三个随机微分方程:

$$d\eta_t = -b\eta_t dt + \sigma dB_t, \qquad dZ_t = Z_t\Phi_t dB_t,$$

$$d\xi_t = \xi_t(b\,dt + \sigma dB_t).$$

随机微分方程(也称随机积分方程)也是一个表达随机过程(特别是连续时间、连续状态的马氏过程)的一个重要而方便的工具.

随机微分方程的一般形式如下:

$$\xi_t(\omega) = \xi_0(\omega) + \int_0^t b(u,\xi_u(\omega))du + \int_0^t \sigma(u,\xi_u(\omega))dB_u(\omega), \tag{9.26}$$

其中第一个积分理解为对任意固定的 $\omega$,对时间参数作普通函数的积分;第二个积分理解为 Ito 随机积分. (9.26)也常写为

$$\begin{cases} d\xi_t = b(t,\xi_t)dt + \sigma(t,\xi_t)dB_t, \\ \xi_0 = \xi_0(\omega). \end{cases} \tag{9.27}$$

上述方程的存在唯一性的研究是随机微分方程的理论研究的一个基本问题,它不仅是随机分析的理论基础,而且对于许多应用问题

也有重要意义.例如,在物理问题中归纳出了 Langevin 方程,从其存在唯一性,就可断定它正是如 1 段中例 9.1 给出的那个 Gauss 马氏过程,从而我们就知道它有许多重要的性质.

下面我们给出随机微分方程的一个比较简略的存在唯一性定理.

**定理 9.2** 如果 $\sigma, b$ 满足:∃C 使

$$|\sigma(t,x)-\sigma(t,y)|+|b(t,x)-b(t,y)|\leqslant C|x-y|,$$

那么方程(9.18)存在唯一解 $\xi_t$,它是下面迭代过程的极限:

$$\begin{cases}\xi_t^{(0)}\triangleq\xi_t,\\ \xi_t^{(n+1)}\triangleq\xi_0+\int_0^t b(s,\xi_t^{(n)})\mathrm{d}s+\int_0^t\sigma(s,\xi_t^{(n)})\mathrm{d}B_s,\end{cases}$$

$$\xi_t\triangleq(p)\lim_{n\to\infty}\xi_t^{(n)}.$$

由此可以看出,$\xi_t$ 只依赖于 $\xi_0$ 及 Brown 运动 $B$ 的过去时间的值.

**注:** 定理 9.2 的条件可以用来作数值模拟或数值解.又若只关心解的存在唯一性,则定理条件可减弱.

利用存在唯一性定理,就可以知道满足随机微分方程(9.23)就一定是如 §9.1 中给出的 Gauss 马氏过程.

## * §9.3 应用于滤波与调制信号的解调

### 1. 调制信号与量测方程

回忆第八章中的定义:随机过程 $\{\zeta_t,t\geqslant0\}$ 称为宽平稳的,如果 $E\zeta_t=m$(常数)且 $E(\zeta_s-m)(\zeta_t-m)$ 只依赖 $t-s$(记成 $R(t-s)$,称为 $\zeta_t$ 的协方差函数).

宽平稳过程不妨假定 $E\zeta_t=0$,此时协方差函数就是相关函数.

类似于宽平稳列的 AR($p$)模型,对于宽平稳过程也有"AR($p$)"模型,此时差分方程应代之以形式微分方程:

$$\zeta_t^{(p)} + \alpha_1\zeta_t^{(p-1)} + \cdots + \alpha_p\zeta_t = \sigma_0 W_t,$$

其中$\xi_t^{(k)}$是$\xi_t$关于$t$的$k$阶导数.注意到在宽平稳列时,左方是差分,而右方是i.i.d.序列,也就是"i.i.d.序列的部分和"的增量(差分).现在$t$连续变化,左边变成各阶微商,右边的"i.i.d.列的部分和"应该变成Brown运动$\{B_t,t\geqslant 0\}$(见第一章),于是右边的$W_t$应该是Brown运动$B_t$对$t$的微商.但是Brown运动的轨道是处处连续处处不可微的(这个事实的证明需要测度论的知识,我们不在本书中列入),微商没有意义,因此上述的形式微分方程应理解为积分方程,就是说,下述Ito方程:

$$d\zeta_t^{(p)} = -(\alpha_1\zeta_t^{(p-1)} + \cdots + \alpha_p\zeta_t)dt + \sigma_0 dB_t,$$

即

$$\zeta_t^{(p)} - \zeta_0^{(p)} = -\int_0^t(\alpha_1\zeta_s^{(p-1)} + \cdots + \alpha_p\zeta_s)ds + \sigma_0 B_t.$$

右端积分中$\zeta_t^{(k)}$是$\zeta_t$关于$t$的$k$次微商:$\zeta_t^{(k)}(\omega) = \dfrac{d\zeta_t^{(k-1)}(\omega)}{dt}(k = 1,2,\cdots,p-1)$,但注意$\zeta_t^{(p)}(\omega)$并不是$\zeta_t^{(p-1)}(\omega)$在通常意义下的微商,只能按上式意义理解.

同宽平稳列类似地,我们可以想象,连续时间的AR($p$)模型在宽平稳过程理论中也称为$p$步马氏模型,它也可以用来近似相当宽的一类宽平稳过程.特别AR($p$)模型还可以看成$p$维的AR(1)模型.

为此我们令

$$\vec{\eta}_t = \begin{pmatrix} \zeta_t \\ \zeta_t^{(1)} \\ \vdots \\ \zeta_t^{(p-1)} \end{pmatrix},$$

则

$$d\vec{\eta}_t = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -\alpha_p & \cdots & & -\alpha_1 \end{pmatrix}\vec{\eta}_t dt + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} dB_t.$$

所以在本节中，我们不妨假定信号为AR(1)模型，也可以是向量型信号. 这里为叙述简单，我们这里还假定信号为一维的：

$$d\zeta_t = -k\zeta_t dt + \sigma_0 dB_t, \tag{9.28}$$

即 $\zeta_t - \zeta_0 = -\int_0^t k\zeta_s ds + \sigma_0 B_t$.

在信号发送时，为了使通路增加容量与抗干扰等的需要，需要进行调制，所谓调制就是实际发送的不是 $\zeta_t$，而是它的一类变换 $g(\zeta_t, t)$. 最常见的调制有以下三种：

| 调制方式： | 调幅 | 调相 | 调频 |
|---|---|---|---|
| $g(\zeta_t, t)$： | $\zeta_t \sin\omega_0 t$ | $\sin(\omega_0 t + \beta\zeta_t)$ | $\sin\left(\omega_0 t + \int_0^t \zeta_s ds\right)$ |

这里的 $\omega_0$ 称为载波频率(圆频率). 收到的信号还混进了噪声，所以量测方程应为

$$d\eta_t = g(\zeta_t, t)dt + \gamma_0 dB_t^{(2)},$$

其中 $\{B_t^{(2)}\}$ 是一个与 $\{B_t\}$ 相互独立的 Brown 运动. 于是在调频的情形的信号方程(**状态方程**)(参见第八章)应改写为：

$$d\begin{pmatrix}\zeta_t \\ \tau_t\end{pmatrix} = \begin{pmatrix}-k & 0 \\ 1 & 0\end{pmatrix}\begin{pmatrix}\zeta_t \\ \tau_t\end{pmatrix}dt + \begin{pmatrix}\sigma_0 \\ 0\end{pmatrix}dB_t^{(1)};$$

**而量测方程**为

$$d\eta_t = \sin(\omega_0 t + \alpha\tau_t)dt + \gamma_0 dB_t^{(2)}.$$

更一般地，在通信系统中常出现的模型可以写成

$$\begin{cases} d\vec{\zeta}_t = A\vec{\zeta}_t dt + D dB_t^{(1)} & (\text{状态方程}) \\ d\eta_t = g(\vec{\zeta}_t, t)dt + \gamma_0 dB_t^{(2)} & (\text{量测方程}). \end{cases} \tag{9.28$'$}$$

其中 $A, D$ 分别为矩阵和向量，$B_t^{(1)}, B_t^{(2)}$ 为相互独立的 Brown 运动(($\eta_t, g(\cdot,\cdot), \gamma_0$)也可以是向量). 这是连续时间参数的滤波问题，由于 $g(x,t)$ 是非线性的，所以这个模型远比 Kalman-Bucy 模型(见第八章)广. 非线性方程(9.28)的解，很难有显式表示. 另一方面，也只能对几个特殊的情形给出非常粗的近似解法，所以(9.28)$'$的模型实际应用起来仍能力有限.

## *2. 最优估计(非线性滤波)

发方信号 $\zeta_t$ 经调制后,经过信号通道,再加上不可避免的噪声到达收方. 收方收到的是量测信号 $\eta_t$. 问题是如何尽可能好地由 $\{\eta_t\}$ 去恢复 $\{\zeta_t\}$? 也就是已知量测过程在 $[0,t]$ 观测到的样本 $\{\eta_s:0\leqslant s\leqslant t\}$,要用它们的某个函数来估计 $\zeta_t$.

假定在 $\{\eta_s,0\leqslant s\leqslant t\}$ 已知的条件下,$\zeta_t$ 的条件密度记为

$$p(t,x\,|\,\eta_s,0\leqslant s\leqslant t).$$

自然 $\zeta_t$ 的估计 $\hat{\zeta}_t$ 应取为 $\{\eta_s,0\leqslant s\leqslant t\}$ 已知条件下,$\zeta_t$ 的条件期望:

$$\hat{\zeta}_t=\int xp(t,x\,|\,\eta_s,0\leqslant s\leqslant t)\mathrm{d}x.$$

这种取法非常直观,并且在数学上可以证明是均方意义下最佳的(见命题 8.2),即均方误差最小的估计. 问题在于怎样求出上式中的条件概率密度.

Kushner 对于非常一般的信号过程(状态方程)得到了上面的条件密度 $p$ 满足的方程,即对于状态方程

$$\mathrm{d}\zeta_t=b(\zeta_t)\mathrm{d}t+\sigma(\zeta_t)\mathrm{d}B_t^{(1)}$$

及量测方程

$$\mathrm{d}\eta_t=g(\zeta_t,t)+\gamma_0\mathrm{d}B_t^{(2)},$$

以上定义的条件分布密度 $p_t=p(t,x\,|\,\eta_s,0\leqslant s\leqslant t)$ 应该满足

$$\mathrm{d}p_t=\left(\frac{1}{2}\frac{\partial^2(\sigma^2(x)p)}{\partial x^2}-\frac{\partial(b(x)p)}{\partial x}\right)\bigg|_{x=\zeta_t}\mathrm{d}t+(g(\zeta_t,t)$$

$$-\hat{g}(\zeta_t,t))\times\frac{p_t}{\gamma_0}(\mathrm{d}\eta_t-\hat{g}(\zeta_t,t)\mathrm{d}t),$$

其中 $\hat{g}(\zeta_t,t)$ 是 $g(\zeta_t,t)$ 关于 $p(t,x\,|\,\eta_s,0\leqslant s\leqslant t)$ 的期望:

$$\hat{g}(\zeta_t,t)=\int g(x,t)p(t,x\,|\,\eta_s,0\leqslant s\leqslant t)\mathrm{d}x.$$

由于在观测前,量测过程 $\eta$ 是随机过程,因此上述关于 $p$ 的方程是随机的偏微分方程(它的推导可参见[Sch]等书),称为 Kushner

方程. 然而 Rushner 方程中尚有$\hat{g}$,一般地它依赖于 $p$,因而一般地 Kushner 方程并未完全提供 $p_t$ 的算法,但在下一段的线性情形,$g$ 是常数,从而$\hat{g}$ 就不依赖 $p$ 了.

对于光滑函数 $f$,由 Kushner 方程可得到对于 $f(\zeta_t)$的最佳估计$\hat{f}(\zeta_t)$,即

$$\hat{f}(\zeta_t) = \int f(x)p(t,x|\eta_s,0 \leqslant s \leqslant t)\mathrm{d}x,$$

应满足方程

$$\begin{aligned}\mathrm{d}\hat{f}(\zeta_t) &= \int f(x)p_t\mathrm{d}x \\ &= \left\{\frac{1}{2}\widehat{\sigma^2 f''} + (\widehat{bf'})\right\}\Bigg|_{x=\zeta_t}\mathrm{d}t \\ &\quad + [(fg)(\zeta_t,t) - \hat{f}(\zeta_t)\hat{g}(\zeta_t,t)] \times \frac{1}{\gamma_0}(\mathrm{d}\eta_t - \hat{g}(\zeta_t,t)\mathrm{d}t).\end{aligned} \tag{9.29}$$

对于 $d$(一般 $d>1$) 维的$\vec{\zeta}_t$ 的情形,设状态方程为

$$\mathrm{d}\vec{\zeta}_t = \vec{b}(\vec{\zeta}_t)\mathrm{d}t + \sigma(\vec{\zeta}_t,t)\mathrm{d}\vec{B}_t^{(1)}$$

($\sigma$ 为矩阵, $\vec{B}_t^{(1)}$,$\vec{b}$ 均为向量). 又设量测过程也是多维的,即

$$\mathrm{d}\vec{\eta}_t = g(\vec{\zeta}_t,t) + \Gamma_t\mathrm{d}\vec{B}_t^{(2)}$$

($g$ 为向量值函数,$\vec{B}_t^{(2)}$为向量,$\Gamma_t$ 为矩阵),那么对多维的光滑函数 $f(\vec{x})$,估计$\hat{f}(\vec{\zeta}_t)$应满足随机微分方程:

$$\begin{aligned}\mathrm{d}\hat{f}(\vec{\zeta}_t) &= (\widehat{Lf})\mathrm{d}t + [(fg)(\vec{\zeta}_t,t) - \hat{f}(\vec{\zeta}_t)\hat{g}(\vec{\zeta}_t,t)] \\ &\quad \times (\Gamma_t\Gamma_t^T)^{-1}(\mathrm{d}\vec{\eta}_t - \hat{g}(\vec{\zeta}_t,t))\mathrm{d}t.\end{aligned} \tag{9.30}$$

此处

$$Lf(x) = \frac{1}{2}\sum_{i,j}(\sigma(x)\sigma(x)^T)_{i,j}\frac{\partial^2 f}{\partial x_i\partial x_j} + b(x)\nabla f.$$

取 $f=x_i$,则

$$\hat{\zeta}_t = \begin{pmatrix} \hat{x}_1(\vec{\zeta}_t) \\ \vdots \\ \hat{x}_d(\vec{\zeta}_t) \end{pmatrix}.$$

记它的平方误差阵为

$$D(\vec{\zeta}_t) \triangleq (\vec{\zeta}_t - \hat{\zeta}_t)(\vec{\zeta}_t - \hat{\zeta}_t)^T. \tag{9.31}$$

将它的最佳估计记为

$$V_t \triangleq \hat{D}_t(\vec{\zeta}_t). \tag{9.32}$$

于是也可以得出 $V_t$ 满足的随机微分方程.

### 3. 线性情形(连续时间参数的 Kalman-Bucy 滤波)

这时我们有

$$\begin{cases} \mathrm{d}\vec{\zeta}_t = A_t\vec{\zeta}_t\mathrm{d}t + D_t\mathrm{d}\vec{B}_t^{(1)} & \text{(状态方程)}, \\ \mathrm{d}\vec{\eta}_t = M_t\vec{\zeta}_t\mathrm{d}t + \Gamma_t\mathrm{d}\vec{B}_t^{(2)} & \text{(量测方程)}, \end{cases} \tag{9.33}$$

$\vec{B}_t^{(1)}, \vec{B}_t^{(2)}$ 为相互独立的多维 Brown 运动，$A_t, D_t, M_t, \Gamma_t$ 为非随机的矩阵.

记 $R_t = \Gamma_t\Gamma_t^T$，由前面的讨论可以得出误差阵的最佳估计 $V_t$ 满足 **Riccatti 方程**(由于推导较长，本书将其略去，读者可参考[Sch])

$$\frac{\mathrm{d}V_t}{\mathrm{d}t} = A_tV_t + VA_t^T + D_tD_t^T - V_tM_t^TR_t^{-1}M_tV_t, \tag{9.34}$$

而 $\hat{\zeta}_t$ 的方程则化为

$$\mathrm{d}\hat{\zeta}_t = A_t\hat{\zeta}_t\mathrm{d}t + V_tM_t^TR_t^{-1}(\mathrm{d}\vec{\eta}_t - M_t\hat{\zeta}_t\mathrm{d}t), \tag{9.35}$$

(9.35)称为 Kalman-Bucy 滤波器，$K_t \triangleq V_tM_t^TR_t^{-1}$ 称为 Kalman 增益，要算出它必须先解 Riccatti 方程(9.34).

**注 1**：这里线性滤波器的形式与离散情形(见第八章)很相似. 其实这里无非是用 $V_t$ 代替了那里的 $P_n$，Riccatti 方程代替了差分式. 然而两者也有

不同之处：在离散情形

$$P_n = E(\vec{x}_n - \hat{x}_n)(\vec{x}_n - \hat{x}_n)^T$$

是滤波的 $\hat{x}_n$ 的理论误差方差阵，而在这里 $V_t$ 是方差阵，

$$(\vec{\xi}_t - \hat{\xi}_t)(\vec{\xi}_t - \hat{\xi}_t)^T$$

的最佳估计(它是估计误差的估计).

**注 2**： 另一个推导 Kalman-Bucy 滤波器的方法是用**新息过程**(参见[Ok])，滤波误差阵则仍用 $\hat{\xi}_t$ 的理论误差方差阵

$$S_t = E(\vec{\xi}_t - \hat{\xi}_t)(\vec{\xi}_t - \hat{\xi}_t)^T,$$

它也满足 Riccatti 方程. 增益 $K_t \triangleq S_t M_t^T R_t^{-1}$. 注意 $S_t = ED(\vec{\xi}_t)$，而 $V_t = \hat{D}(\vec{\xi}_t)$ 它们的定义是不同的，但是它们满足同一个 Riccatti 方程，所以只要 $S_0 = V_0$，它们的解也有 $S_t = V_t$，从而得到的估计 $\hat{\xi}_t$ 就是一样的. 这说明在线性情形，用 $S_t$ 与用 $V_t$ 差别不大.

一般地，如果 $\vec{\xi}_0$ 已知，则取 $\hat{\xi}_0 = \vec{\xi}_0, S_0 = V_0 = \mathrm{Var}\,\vec{\xi}_0$，如果 $\vec{\xi}_0$ 未知，则常可取 $\hat{\xi}_0 = 0, S_0 = V_0 = 0$.

**注 3**： 又如果 Riccatti 方程的解有一个极限 $V^{(\infty)}(\triangleq \lim_{t\to\infty} V_t)$，则可以取 $S_0 = V_0 = V^{(\infty)}$，此时 $S_t = V_t = V^{(\infty)}$ 与 $t$ 无关. 在 $A_t, D_t, M_t, \Gamma_t$ 等均为不依赖 $t$ 的常矩阵时，常可能出现此种情形，此时 $V$ 是代数方程

$$AV + VA^T + DD^T - VM^TR^{-1}MV = 0$$

的非负矩阵解.

**注 4**： 一维情形，即 $A_t, D_t, M_t, \Gamma_t$ 等都是数值函数时，则滤波化为

$$\begin{cases} \dfrac{\mathrm{d}V_t}{\mathrm{d}t} = 2A_tV_t - \dfrac{M_t^2V_t^2}{R_t} + B_t^2, \\ \mathrm{d}\hat{\xi}_t = \left(A_t - \dfrac{M_t^2V_t}{R_t}\right)\hat{\xi}_t\mathrm{d}t + \dfrac{M_tV_t}{R_t}\mathrm{d}\eta_t. \end{cases}$$

上述方程可以用下述迭代方程

$$\begin{cases} \dfrac{\mathrm{d}V_t^{(n+1)}}{\mathrm{d}t} = V_0 + \displaystyle\int_0^t \left(2A_sV_s - \dfrac{M_s^2(V_s^{(n)})^2}{R_s} + B_s^2\right)\mathrm{d}s, \\ \hat{\xi}^{(n+1)}_t = \hat{\xi}_0 + \displaystyle\int_0^t \left(A_s - \dfrac{M_s^2V_s^{(n)}}{R_s}\right)\hat{\xi}_s\mathrm{d}s + \int_0^t \dfrac{M_sV_s^{(n)}}{R_s}\mathrm{d}\eta_s \end{cases}$$

作近似计算. 特别在一维且 $A_t, D_t, M_t, \Gamma_t$ 等不依赖 $t$ 时，$V^{(\infty)}$(如果存在)是二

次方程

$$2AV - \frac{M^2V^2}{R} + B^2 = 0$$

的正解.

**4. 调制信号的滤波与解调**

在调制信号中,事实上常常隐含调制前的原信号 $\zeta_t$ 的频率大大低于调制频率 $\omega_0$,所以在解调器内常设有低通滤波器,使高频项不能通过.在设计接收信号的近似最佳滤波器的时候,常忽略高频项.低通滤波器有标准的设计与产品(硬件或软件),在本书中不再叙述.

1° 经过低通的调辐传输信号的 Kalman-Bucy 滤波器.

设信号方程为:

$$d\zeta_t = - k\zeta_t dt + b dB_t^{(1)},$$

而量测方程是调幅信号加上通路噪声:

$$d\eta_t = \zeta_t \sin\omega_0 t dt + \gamma dB_t^{(2)} \quad (B^{(1)},\ B^{(2)} \text{ 独立}).$$

这是一个线性系统.由第 3 段(9.34),误差方差阵应满足

$$\dot{V}_t = - 2kV_t - 2\frac{\sin^2\omega_0 t}{\gamma^2}V_t^2 + b^2,$$

且 $$R_t = \gamma^2, \quad M_t = \sin\omega_0 t, A_t = - k.$$

注意到

$$\sin^2\omega_0 t = \frac{1}{2}(1 - \cos 2\omega_0 t),$$

在经过低通滤波器后,滤去高频项 $\cos 2\omega_0 t$,只留下了 $\frac{1}{2}$,于是

$$\dot{V}_t \approx - 2kV_t - \frac{1}{2\gamma^2}V_t^2 + b^2.$$

具体解出这个常微分方程,可以看到这个渐近方程的近似稳态方程为

$$- 2kV_t - \frac{1}{2\gamma^2}V_t^2 + b^2 = 0,$$

其解为

$$V^{(\infty)} = (4k^2\gamma^4 + 2\gamma^2 b^2)^{\frac{1}{2}} - 2k\gamma^2.$$

取 $V_0 = V^{(\infty)}$，则 $V_t \approx V^{(\infty)}$，由(9.35)就得到$\hat{\zeta}_t$的近似方程

$$\mathrm{d}\hat{\zeta}_t = -\left\{k\hat{\zeta}_t \mathrm{d}t + \frac{V^{(\infty)}}{\gamma^2}\sin\omega_0 t d\eta_t\right\} + \frac{V^{(\infty)}}{\gamma^2}\sin^2\omega_0 t \mathrm{d}t. \tag{9.36}$$

同样的理由，在 $\sin^2\omega_0 t$ 中由低通滤波器滤去高频项后只留下了$\frac{1}{2}$，于是最后得 Kalman-Bucy 滤波器近似地由下面的方程描述：

$$\mathrm{d}\hat{\zeta}_t = -\left\{k + \frac{V^{(\infty)}}{2\gamma^2}\right\}\hat{\zeta}_t \mathrm{d}t + \frac{V^{(\infty)}}{\gamma^2}\sin\omega_0 t \mathrm{d}\eta_t.$$

从而有显式解

$$\hat{\zeta}_t = \hat{\zeta}_0 \mathrm{e}^{-\left(k + \frac{V^{(\infty)}}{2\gamma^2}\right)t} + \frac{V^{(\infty)}}{\gamma^2}\int_0^t e^{-\left(k+\frac{V^{(\infty)}}{2\gamma^2}\right)(t-s)}\sin(\omega_0 s)\mathrm{d}\eta_s.$$

2° 经过低通滤波器的调相传输信号 Kalman-Bucy 滤波器的粗近似为：

信号方程仍为

$$\mathrm{d}\zeta_t = -k\zeta_t \mathrm{d}t + b\mathrm{d}B_t^{(1)},$$

但是量测方程是调相信号加上通路噪声，即：

$$\mathrm{d}\eta_t = \sin(\omega_0 t + \beta\zeta_t) + \gamma \mathrm{d}B_t^{(2)} \quad (B^{(1)}, B^{(2)} \text{ 独立}).$$

这时量测方程是非线性的，我们要给出滤波方程(9.35)及估计方差 $V_t$ 满足的方程(9.34)的线性粗近似. 这里所谓粗近似是指：推导近似方程时**需要假定** $\zeta_t - \hat{\zeta}_t$ 很小，但是在实际应用中并不能保证$\zeta_t - \hat{\zeta}_t$很小，所以这种近似方程是非常粗的. 以下我们只列出其梗概，而略去一切直观推导.

把(9.29)的右端按$\zeta_t - \hat{\zeta}_t$ 作 Taylor 展开只留下低阶项；并考虑到$\zeta_t - \hat{\zeta}_t$的估计值为 0，在用 $V_t$ 代替 $EV_t$ 后就得到以下的粗近似方程

$$d\hat{\zeta}_t = -k\hat{\zeta}_t dt + \frac{1}{\gamma^2}\frac{\partial g}{\partial x}(\hat{\zeta}_t, t)V_t(d\eta_t - g(\hat{\zeta}_t, t)dt).$$

对于估计方差 $V_t$ 的方程，作同样的粗近似后也得到粗近似方程

$$dV_t \doteq (b^2 - 2kV_t)dt + \frac{1}{\gamma^2}V_t^2\frac{\partial^2 g}{\partial x^2}(\hat{\zeta}_t, t)V_t(d\eta_t - g(\hat{\zeta}_t, t)dt).$$

利用 $g(x, t) = \sin(\omega_0 t + \beta x)$ 代入得

$$dV_t \doteq (b^2 - 2kV_t)dt + \frac{\beta^2 V_t^2}{\gamma^2}[(-\sin(\omega_0 t + \beta\zeta_t)\sin(\omega_0 t + \beta\hat{\zeta}_t)$$

$$+ \sin^2(\omega_0 t + \beta\hat{\zeta}_t))dt - \gamma\sin(\omega_0 t + \beta\hat{\zeta}_t)dB_t^{(2)}],$$

再注意

$$\sin(\omega_0 t + \beta\zeta_t)\sin(\omega_0 t + \beta\hat{\zeta}_t)$$

$$= \frac{1}{2}[\cos(2\omega_0 t + \beta(\zeta_t + \hat{\zeta}_t)) + \cos\beta(\hat{\zeta}_t - \zeta_t)]$$

及

$$\sin^2(\omega_0 t + \beta\hat{\zeta}_t) = \frac{1}{2}(1 - \cos(2\omega_0 t + 2\beta\hat{\zeta}_t)).$$

由低通滤波器滤去高频部分（含 $2\omega_0 t + *$ 的部分），并利用“$\zeta_t - \hat{\zeta}_t$ 小时 $\cos\beta(\zeta_t - \hat{\zeta}_t) \approx 1$”，再假定 $\gamma \ll 1$，于是 $V_t$ 的最后粗近似方程为 $dV_t \doteq b^2 - 2kV_t$．此方程有稳定解 $V_t = \frac{b^2}{2k}$ $\left(取 V_0 = \frac{b^2}{2k}，则 V_t = \frac{b^2}{2k}\right)$.

对于 $\hat{\zeta}_t$，再注意到

$$\frac{\partial g}{\partial x}g(\hat{\zeta}_t, t) = \beta\cos(\omega_0 t + \beta\hat{\zeta}_t)\sin(\omega_0 t + \beta\zeta_t)$$

$$= \frac{\beta}{2}\sin(2\omega_0 t + 2\beta\hat{\zeta}_t).$$

它是高频项，通过低通滤波器后得到 0，于是 $\hat{\zeta}_t$ 的粗近似方程化为最后的形式：

$$\mathrm{d}\hat{\zeta}_t = -k\hat{\zeta}_t\mathrm{d}t - \frac{b^2\beta}{2k\gamma^2}\cos(\omega_0 t + \beta\hat{\zeta}_t)\mathrm{d}\eta_t.$$

3° 调相传输信号的粗近似滤波器的获得比较复杂，有兴趣的读者可参见：

Snyder, D. L., The state-variable approach to continuous estimation withapplication to analog communication Theory, MIT Press, Cambridge, Mass, 1969.

## § 9.4 证券投资模型与随机控制

### 1. 问题的提出

我们先举一例.

**例 9.7**(证券的最佳投资方案)

有价证券可以分为两种. 一种是不担风险的，如银行存款，短期国债等. 另一种是有风险的(但可能收益高的)，如股票等. 今某人有资本 $x$ 元，在不考虑借贷的情形下，问在每个时刻 $t$(连续时间)以什么比例购买这两种证券，可使得他在时间 $T$ 的时候获益最多？

在时刻 $t=0$ 时的资金 $x$，把它用于购买这两种证券. 为简单起见，我们不计卖出买进所需的时间、交易费与税金. 设在任意 $t<T$，他分别以 $u_t$ 与 $1-u_t$ 的比例($0\leqslant u_t\leqslant 1$)买进有风险与无风险证券. 这样使他在 $t$ 时刻的总财富为 $\xi_t(\xi_0=x)$. 设无风险证券的利率为 $r(>0)$，有风险 证券的利率为 $b(>r)$ 它还具有随机波动风险，假定这种随机波动风险可以用 Brown 运动来表达，于是我们可以采用下述随机微分方程的模型来描述 $\xi_t$(风险证券就是 § 9.2 节的 Black-Scholes 模型)：

$$\begin{cases} d\xi_s = u_s\xi_s(b\mathrm{d}s + \sigma\mathrm{d}B_s) + (1-u_s)\xi_s r\mathrm{d}s \quad (s>0), \\ \xi_0 = x \quad (0\leqslant u_s\leqslant 1), \end{cases}$$

其中 $r, b, \sigma$ 都是非负常数，$b>r$，$\{B_t: t\geqslant 0\}$ 是 Brown 运动.

我们的目的是想"调整"$u_t$,使它在某个固定时刻 $T$ 有最大的平均财富,即 $E\xi_T=$ 最大(更一般地,对某个目标函数 $\varphi(x)$,能够使 $E\varphi(\xi_T)=$ 最大).

注意到 $\sigma$ 可以理解为一种"风险随机波动率",当 $B_t$ 取绝对值很大的负值时,此人就要大损失,这里就是一个随机过程,如果 $\xi_t$ 首达 0 的随机时刻为 $\tau_0$,那么购买过程到 $\tau_0$ 必须终止,因为此时他已经把财富耗尽了. 所以我们的目标应修改为:要选取 $u_t$,使 $E\xi_{T\wedge\tau_0}$ 最大.

对于 $u_t$,除了取值在[0,1]内的限制以外,它应依赖于随机过程 $\{\xi_t: t\geqslant 0\}$,而且只依赖于 $\xi$ 的过去的值(也就是不依赖于 $\xi$ 的将来的值).

这类问题与 § 4.5 的随机决策类似,不同之处在于这里研究的 $\xi_t$ 属于连续时间连续状态的情况. 这时我们有一个由随机微分方程决定的随机系统,其中包含一个可控制的因素 $u$.

**2. 随机控制的模型**

为了能研究多维的情形,我们要简要地叙述关于多维 Brown 运动的 Ito 随机积分的定义. 设 $\{B_t: t\geqslant 0\}$ 是一个 $m$ 维 Brown 运动,即它是 $m$ 个相互独立的 Brown 运动 $\{B^{(i)}: i\leqslant m\}$ 组成的列向量. 对于 $d\times m$ 矩阵值随机函数 $\{(\Phi_{ij}(t,w))_{i\leqslant d}: t\geqslant 0\}$,只要满足:

(1) $\Phi_{ij}(t,w)$ 只依赖 $B$ 过去的值;

(2) $\int_a^b E|\Phi_{ij}(t,w)|^2\mathrm{d}t<\infty$,

那么我们定义

$$\int_a^b \Phi_t \mathrm{d}B_t \triangleq \begin{pmatrix} \sum_1^m \int_a^b \Phi_{1j}(t,w)\mathrm{d}B_t^{(j)} \\ \vdots \\ \sum_1^m \int_a^b \Phi_{dj}(t,w)\mathrm{d}B_t^{(j)} \end{pmatrix}.$$

现在我们考虑 $d$ 维随机微分方程，其中带有一个 $k$ 维可控制因素 $u$.

$$\begin{cases} d\xi_s = \sigma(\xi_s, u_s)dB_s + b(\xi_s, u_s)ds \quad (s \geqslant t), \\ \xi_t = x, \end{cases} \tag{9.36}$$

其中 $\xi$ 是 $d$ 维随机过程，$u$ 是只依赖 $\xi$ 的过去的值的 $k$ 维随机过程，$B$ 是 $m$ 维 Brown 运动，$\sigma, b$ 分别为 $d\times m, d\times 1$ 矩阵值函数. $u$ 代表控制策略或控制方案. 一般讲 $u$ 是一个随机过程，它决定在每个时刻 $t$ 用来控制 $\xi$ 的控制量 $u_t$；它由系统在 $t$ 以前的情况来决定，因此 $u$ 不能依赖于 $\xi$ 的将来，而是只依 $\xi$ 的过去的值的随机过程，也就是只依赖 $\xi_0$ 与 $B$ 的过去的值的过程. 一旦 $u$ 依赖于 $\xi$ 的方式选定(即给定了一个控制方案或控制策略后)，上述方程的解过程 $\xi$ 是依赖于 $u$ 的，故也常记为 $\xi^u$. 控制的目的是要选取合适的 $u$，使得某个与 $\xi$ 有关的量达到最优，例如说使某种代价达最小、或使某种报酬达最大.

为了数学上的要求，例如说使方程(9.36)存在唯一的解 $\xi$，我们假定 $b(x,v)$ 与 $\sigma(x,v)$ 十分好，例如它们都满足 Lipschitz 条件，甚至有一定的可微性. 我们把这唯一的解记为 $\xi_s^{t,x}$(其中上标表示初值 $\xi_t = x$，此外它也依赖 $u$. 在不引起记号混淆下，我们略去了这个解关于 $u$ 依赖性的足标).

假定在时刻 $s$ 系统处于状态 $x$，且控制策略的值为 $v$ 时，单位时间的运行代价为 $F(s,x,v)$，而在时刻 $s$ 停止且系统处于状态 $x$ 时的停止操作代价为 $\varphi(s,x)$，于是随机系统 $\xi^{t,x}$ 在控制策略(过程)$\{u_s: t\leqslant s\leqslant \tau\}$ 下的平均代价为

$$J(t,x,u_\cdot) = E\left[\int_t^\tau F(s,\xi_s^{t,x},u_s)ds + \varphi(\tau,\xi_\tau^{t,x})\right]. \tag{9.37}$$

这里的 $\tau$ 通常是 $\xi$ 越出某个区域的时间，它是一个停时，即对任意 $t\geqslant 0$，$\{\tau\leqslant t\}$ 总是过程 $\xi$ 在 $t$ 以前发生的一件事.

记全体可能采取的控制策略集合为 $U$，而记

$$U_M \triangleq \{u(t,\xi_t): 任意二元函数 u(t,x)\},$$

$$U_S \triangleq \{u(\xi_t)：任意一元函数 u\}.$$

显见 $U_S \subset U_M \subset U$. $U_M$ 中的控制策略叫做**马氏控制策略**. 因为在时刻 $t$ 的策略 $U_t$ 只依赖于系统在时刻 $t$ 的状况 $\xi_t$, 而与系统在时刻 $t$ 以前的状况无关;$U_S$ 中的控制策略称为**平稳马氏控制策略**. 因为这种策略只依赖于系统状况,不随时间改变这种依赖关系. 非随机的控制策略称为**决定性的控制**或**开环控制**. 随机的控制策略称为**闭环控制**或**反馈控制**.

最佳控制的基本问题是:

1° 如何求 $V(t,x) \triangleq \inf\limits_{u\in U} J(t,x,u)$(最佳代价)?

2° 是否存在(是否唯一)$u^* \in U$,使

$$J(t,x,u^*) = \inf_{u\in U} J(t,x,u) \text{(}u^* \text{ 称为最佳控制策略)?}$$

3° 若上述 $u^*$ 存在,则如何找出它?

4° 若 $u^*$ 不存在,则如何求 $\varepsilon$-最佳控制策略 $u^\varepsilon$,使

$$J(t,x,u^\varepsilon) < \inf_{u\in U} J(t,x,u) + \varepsilon?$$

5° 什么时候有

$$\inf_{u\in U_M} J(t,x,u) = \inf_{u\in U} J(t,x,u),$$

即什么时候存在最佳策略是马氏控制策略?如果它存在,则又如何去找到它?

**注 1**: 如果在控制的过程中不能量测到 $\xi_t^{(u)}$ 的值,而只能量测到如下的过程 $\eta$:

$$d\eta_s = h(s,\xi_s^{(u)})ds + \gamma(s,\xi_s^{(u)})d\tilde{B}_s,$$

其中 $\tilde{B}$ 是与 $B$ 相互独立的 Brown 运动,那么容许的控制策略 $u$ 不应依赖的 $\xi^{(u)}$ 过去值,而应依赖 $\eta$ 的过去值,这时我们遇到的就不再是一个单纯的控制问题,而是带有滤波的控制问题.

在非常特殊的情形,例如 $\sigma$ 只依赖于 $x$(即 $\sigma(x,v)=\sigma(x)$ 的情形),且 $\sigma$ 与 $b$ 都是线性函数,而 $F,\varphi$ 都是二次函数的情形,人们早就证明了滤波与控制可以分开地去做,也就是在带有量测方程的控制问题中,先假定没有控制(置 $u=0$)求得过程 $\xi$ 的滤波过程 $\hat{\xi}$;再假定 $\xi$ 可无误差地量测(即没有附加的量测方程的情形),由此求得最佳控制策略;最后在求得的最佳控制策略

中用$\hat{\xi}$代替$\xi$. 这样就得到了带有量测的最佳控制. 这种分开的做法称为**分离性原则**. 如果由分离性原则得到的控制策略确是最佳控制策略，则称**分离性原则成立**.

**注 2**: 非时齐(即系数显含$s$的情形)状态方程情形

非时齐情形的受控制状态方程应为

$$\begin{cases} d\xi_s = \sigma(s,\xi_s,u_s)dB_s + b(s,\xi_s,u_s)ds \quad (s \geqslant t), \\ \xi_t = x, \end{cases}$$

这时，只要我们把状态变量$\xi_s$改成新的状态变量$\bar{\xi}_s \triangleq (\xi_s, s)$ ($d+1$维)，就可以得到一个关于$\bar{\xi}_s$的时齐的(即系数不显含$s$)状态控制方程，为此我们记

$$\bar{b}(\bar{\xi}_s,u) \triangleq b(s,\xi_s,u), \quad \bar{\sigma}(\bar{\xi}_s,u) \triangleq \sigma(s,\xi_s,u),$$

于是就有

$$\begin{cases} d\bar{\xi}_s = \begin{pmatrix} \bar{\sigma}(\bar{\xi}_s,u_s) \\ 0 \end{pmatrix} dB_s + \begin{pmatrix} \bar{b}(\bar{\xi}_s,u_s) \\ 1 \end{pmatrix} ds, \\ \bar{\xi}_t = \begin{pmatrix} x \\ t \end{pmatrix}, \end{cases} \tag{9.38}$$

这就变成了时齐的模型.

## *3. 三个典型的最佳控制的定理

1) 单位时间运行代价为时齐情形时越出某区域$D$前的最佳控制:

这时停止时刻$\tau$取为在时刻$t$以后越出$D$的时刻:

$$\tau \triangleq \inf\{s \geqslant t : \xi_s \overline{\in} D\},$$

设$t=0$，并设价格函数为

$$J(x,u) = E\left( \int_0^\tau F(\xi_s,u_s)ds + \varphi(\xi_\tau) \right),$$

其中$F,\varphi$为有界连续函数. 假定$D$为边界光滑的(例如$C^2$边界，即边界上每一点附近的边界都可以用二次可微函数表达的情形)有界区域，而

$$L \triangleq \frac{1}{2}\sum_{i,j=1}^d a_{ij}\frac{\partial^2}{\partial x_i \partial x_j} + b\nabla \tag{9.39}$$

($\nabla$是梯度算子，$a_{ij}=(\sigma\sigma')_{ij}$).

**定理 9A** 在上述条件下，我们有：

假定在马氏控制策略中存在不显含 $t$ 的最佳的马氏控制策略 $u^*$(存在某个函数 $f$，使 $u_t^*=f(\xi_t)$)，即它满足

$$V(x)\triangleq\inf_{u_t\in\Gamma_M}J(x,u_\cdot)=J(x,u^*),\qquad(9.40)$$

再假定 $V(x)$二次连续可微，则 $u^*$ 也是最佳控制策略（从而 $V(x)$ 为最低价格函数）．而且 $V(x)$满足下述椭圆型 HJB（Hamilton-Jacobi-Bellman)极值微分方程的如下边值问题：

$$\begin{cases}\inf\limits_{v\in R^k}(F(x,v)+(LV)(x,v))=0, & x\in D,\\ V|_{\partial D}=\varphi.\end{cases}$$

反之，如果存在在 $D$ 的闭包$\bar{D}$上连续且在 $D$ 二次连续可微的有界函数$W$满足

$$\begin{cases}F(x,v)+(LW)(x,v))=0\quad(\forall x\in D,\forall v\in R^k),\\ W|_{\partial D}=\varphi.\end{cases}$$

那么 $W\leqslant\inf\limits_{u\in\Gamma_M}J(t,u)$. 如果进一步还存在 $D$ 上函数 $f$ 满足

$$F(x,f(x))+(LW)(x,f(x)))=0\quad(\forall x\in D).\ (9.41)$$

那么 $f(\xi)$是最佳控制策略，而 $W(x)$就是最低价格函数.

**注**：这个定理同时也给出了不显含 $t$ 的最佳马氏控制策略 $u^*$ 是最佳控制策略的充分条件，即

$$J(x,u^*)\in C^2.$$

2) 单位时间运行代价(假定依赖时间)在定长时间间隔中的最佳控制：

对于固定时刻 $t_1>t$，记

$$J(t,x,u)=E\left[\int_t^{t_1}F(s,\xi_s^{t,x},u_s)\mathrm{d}s+\varphi(\xi_{t_1}^{t,x})\right],$$

要求出

$$V(t,x)=\inf_{u\in\Gamma_M}J(t,x,u).$$

**定理 9B** 在与定理 9A 相似的条件下，为求最佳控制策略

$u^*$，只需求解抛物性 HJB 方程终值问题：

$$\begin{cases}\inf\limits_{v\in R^k}\left[F(t,x,v)+\dfrac{\partial V}{\partial t}+(LV)\right]=0,\\ V(t_1,x)=\varphi(x)\end{cases}$$

的解 $V(t,x)$，以及寻找 $f(t,x)$，使

$$J(t,\ x,\ f(t,\xi_t))=V(t,\ x),$$

这时 $u_t^*=f(t,\xi_t)$.

3) 混合情形：

设 $\tau$ 是 $\xi$ 在时刻 $t$ 以后越出 $D$ 的时刻，$D$ 满足 1)中条件，固定时刻 $t_1>t$. 又设

$$J(t,x,u)=E\left[\int_t^{t_1\wedge\tau}F(s,\xi_s^{t,x},u_s)\mathrm{d}s+\varphi(\xi_{t_1\wedge\tau}^{t,x})\right],$$

$$V(t,x)=\inf_{u\in U}J(t,x,u).$$

对这个模型，我们有

**定理 9C** 在与定理 9A 相似的条件下，为求最佳控制策略 $u^*$，只需求解抛物性 HJB 方程的终值边值问题：

$$\begin{cases}\dfrac{\partial V}{\partial t}\inf\limits_{v\in R^k}(F(t,x,v)+(LV))=0,\\ V(t_1,x)=\varphi(x),V(t,x)|_{x\in\partial D}=\varphi(x)\quad(t\leqslant t_1)\end{cases}$$

的解 $V(t,x)$，以及寻找 $f(t,x)$，使

$$J(t,x,f(t,\xi_t))=V(t,x).$$

这时 $u_t^*=f(t,\xi_t)$（参见：Bernt Oksendal, Stochaastic Differential Equations, 2nd edition, 1989).

## 4. 线性调整器

设

$$\begin{cases}\mathrm{d}\xi_s=(A_s\xi_s+E_su_s)\mathrm{d}s+\sigma_s\mathrm{d}B_s)\quad(s\geqslant t),\\ \xi_t=x,\end{cases}\tag{9.42}$$

$$J(t,x,u)=E\left[\int_t^{t_1}(\xi_s^TC_s\xi_s+u_s^TD_su_s)\mathrm{d}s+\xi_{t_1}^T\Phi\xi_{t_1}\right],\tag{9.43}$$

其中$\sigma_t$是不依赖于$u_t$的非随机函数，$x$为$d$维实向量，$u$为$k$维实向量，系数均为非随机的时间函数，$A_t$、$C_t$、$\Phi$：$d\times d$，$E_t$：$d\times k$，$D_t$：$k\times k$，$C_t$及$\Phi$为非负定，$D_t$为正定.

此时的 HJB 方程为

$$\begin{cases}\dfrac{\partial V}{\partial t}+x^tC_tx+\inf\limits_{v\in R^k}(v^tD_tv+E_tv\cdot\mathrm{grad}V)\\ \qquad -A_tx\cdot\mathrm{grad}V+\dfrac{1}{2}\sum\limits_{i,j=1}^{d}a_{ij}\dfrac{\partial^2V}{\partial x_i\partial x_j}=0.\\ V(t_1,x)=x^t\Phi x,\end{cases}\tag{9.44}$$

其中$(a_{ij})\triangleq(\sigma\sigma^t)_{ij}$.

我们用待定系数法来试探下述形式的解的可能性：

$$V(t,x)=x^TS_tx+\alpha_t.$$

其中$S_t$非负定，$S_t$与$\alpha_t$都是非随机的时间函数.

由“终端条件”$V(t,x)=x^T\Phi x$显见，我们有：$S_{t_1}=\Phi$，$\alpha_{t_1}=0$.

将上面$V(t,x)$的特殊形式代入抛物型 HJB 方程(9.44)中被取最小的部分，得到

$$\begin{aligned}v^tD_tv&+(E_tv)^t(S_tx+S_t^Tx)\\ &=(v+D_t^{-1}E_t^TS_tx)^TD_t(v+D_t^{-1}E_t^TS_tx)\\ &\quad-(E_t^TS_tx)^TD_t^{-1}(E_t^TS_tx).\end{aligned}$$

可见，取

$$v=-D_t^{-1}E_t^TS_tx,$$

(9.44)左方就达到最小. 把它与$V(t,x)=x^TS_tx+\alpha_t$代入抛物型 HJB 方程(9.44)就变为

$$\begin{cases}\dot S_t=-2A_t^tS_t+S_tE_tD_t^{-1}E_t^TS_t^T-C_t,\\ \dot\alpha_t=-tr(\sigma\sigma^TS_t)\quad(t<t_1).\end{cases}\tag{9.45}$$

我们在终端条件$S_{t_1}=\Phi$下先解(9.45)中第一个方程(它是$S_t$的 Riccatti 方程)，便得到$S_t$. 再解第二个方程并由此用定理 9B 便可得到最低价格为

$$V(t,x) = x^T S_t x + \int_t^{t_1} [tr(\sigma\sigma^T)_s] S_s ds \quad (t < t_1).$$

而最佳控制策略就应是

$$u^* = - D_t^{-1} E_t^T S_t \xi_t,$$

它是一个马氏控制策略.

**注 1**: 如果 $\xi$ 不能直接量测,而量测量是由下述方程确定的 $\eta$:

$$\begin{cases} d\eta_s = H_s \xi_s ds + \Gamma_s d\bar{B}_s, \\ \eta_t = y \quad (s > t), \end{cases} \tag{9.46}$$

其中 $\bar{B}$ 是与 $B$ 独立的 Brown 运动,$H_s$、$\Gamma$ 是非随机的时间函数,那么从线性控制的一般理论中知道"分离原则"是成立的(在本节中我们不再详述"分离原则"在线性控制中成立的理由),所以我们可以先在状态的控制方程中令 $u = 0$,再由量测方程的量测量 $\eta$ 得到 $\xi$ 的滤波值(Kalman Bucy 滤波)$\hat{\xi}_t$,再由上面方法求得无量测情形时($H_s \equiv 1, \Gamma_s \equiv 0$)的最佳控制策略 $-D_t^{-1} E_t^T S_t \xi_t$,最后把 $\xi_t$ 改为 $\hat{\xi}_t$ 便得到最佳控制策略为

$$u^* = - D_t^{-1} E_t^T S_t \hat{\xi}_t.$$

这时由于 $\hat{\xi}_t$ 依赖于 $\eta$ 的过去的取值,从而 $u^*$ 不再是马氏控制策略了(这里的马氏控制策略应理解成$\{u: u_t = f(t, \eta_t)$,任意函数 $f\}$ 中的一个元素).

**注 2**: 在量测方程中不允许出现控制量 $u$.

**注 3**: 求解抛物型 HJB 方程或椭圆型 HJB 方程也是很困难的. 于是人们从更为实用的角度,提出了控制问题的马氏链近似方法(参见:Harold J. Kushner, Paul G. Dupuis, Numerical method for stochastic control problems in continuous time, Springer, 1992).

## 5. 例 9.1 中证券最佳购买方案的确定

现在我们来解决第一段中提出的问题. 我们注意到最大获益问题与最低价格问题之间是对称的,因而在 2～4 的讨论中只要把所有的 inf 都改为 sup,我们都得到了最大获益问题的相应结论.

假定在例 9.1 中的 $\varphi(x) = x^\alpha (0 < \alpha < 1)$(合理的假设应为 $\alpha = 1$,但是这种情形不易求解,故只好假定 $\alpha < 1$),令

$$V(t,x)=\sup_{v\in\Gamma_M} E(\xi^v_{t_1}\wedge\tau_*).$$

这时与3.3)唯一不同之处在于这里的区域 $D=(0,\infty)$ 并不有界，但是事实上在 $d=1$ 的情形下，定理9C仍旧成立. 下面我们试解定理9C中抛物型HJB方程的终值边值问题：

$$\sup_{0\leqslant v\leqslant 1}\left\{\frac{\partial V}{\partial t}+\frac{1}{2}v^2x^2\sigma^2\frac{\partial^2 V}{\partial x^2}+[bv+r(1-v)]x\frac{\partial V}{\partial x}\right\}=0,$$

$$V(t_1,x)=x^a, V(t,0)=0\quad(t\leqslant t_1). \tag{9.47}$$

(9.47)左边等于

$$\frac{\partial V}{\partial t}+\sup_{0\leqslant v\leqslant 1}\left[\left(\frac{1}{2}x^2\sigma^2\frac{\partial^2 V}{\partial x^2}\right)v^2+\left(\frac{2(b-r)\dfrac{\partial V}{\partial x}}{-x\sigma^2\dfrac{\partial^2 V}{\partial x^2}}\right)v\right]+rx\frac{\partial V}{\partial x}.$$

要使方括号内达到最大，只有几种可能：第一种是 $\frac{\partial^2 V}{\partial x^2}>0$，此时有：$v=0$ 时达最大，这样(9.47)变为

$$\begin{cases}\dfrac{\partial V}{\partial t}+rx\dfrac{\partial V}{\partial x}=0,\\ V(t_1,x)=x^a, V(t,0)=0\quad(t\leqslant t_1).\end{cases}$$

第二种是 $\frac{\partial^2 V}{\partial x^2}$ 可正可负，这种情形较为复杂. 由于我们只需要求一个正解，所以不去讨论它. 第三种是 $\frac{\partial^2 V}{\partial x^2}<0$，此时方括号内项在

$$v^*=-(b-r)\frac{\partial V}{\partial x}\Big/x\sigma^2\frac{\partial^2 V}{\partial x^2}\quad(\text{记成 } f(t,x))$$

处取值最大，又由于我们要求 $0\leqslant v\leqslant 1$，从 $v^*\geqslant 0$ 及 $b>r$ 我们看出要使(9.47)有解就必须要求 $\frac{\partial V}{\partial x}\geqslant 0$，于是(9.47)中第一式变为 $\frac{\partial V}{\partial t}+\frac{1}{2}(v^*)^2x^2\sigma^2\frac{\partial^2 V}{\partial x^2}+(r+(b-r)v^*)x\frac{\partial V}{\partial x}=0$. 由 $v^*$ 的表达式我们推出(9.47)等价于

$$
\begin{cases}
\dfrac{\partial V}{\partial t}+rx\dfrac{\partial V}{\partial x}-\dfrac{(b-r)^2\left|\dfrac{\partial V}{\partial x}\right|^2}{2\sigma^2\dfrac{\partial^2 V}{\partial x^2}}=0, & (9.47)' \\
V(t_1,x)=x^{\alpha},\ V(t,0)=0.
\end{cases}
$$

这是一个非线性方程终边值问题.我们试探方程有下述形式的解的可能性:

$$V(t,x)=h(t)x^{\beta}.$$

把它代入(9.47)',便得

$$\beta=\alpha,\ h(t)=e^{\left(r\alpha-\frac{(b-r)^2\alpha}{2\sigma^2(1-\alpha)}\right)(t_1-t)}\xlongequal{\text{记成}}e^{\lambda(t_1-t)}\quad(0<\alpha<1).$$

即

$$V(t,x)=e^{\lambda(t_1-t)}x^{\alpha}.$$

从而

$$f(t,x)=\frac{-(b-r)\dfrac{\partial V}{\partial x}}{x\sigma^2\partial^2 V/\partial x^2}=\frac{b-r}{\sigma^2(1-\alpha)}.$$

由定理 9C,这说明只要$\dfrac{b-r}{\sigma^2(1-\alpha)}<1$ 控制量 $u_t$ 就满足 $0<u_t<1$,而且最佳马氏控制就是常值控制:

$$u_t^*\equiv(b-r)/\sigma^2(1-\alpha).$$

此时$V$满足以上一切要求$\left(在\ 0<x<\infty\ 中有\dfrac{\partial^2 V}{\partial x^2}<0,\dfrac{\partial V}{\partial x}>0\right)$,控制量 $u_t^*$ 满足 $0<u_t^*<1$.这就是:若时刻 $t$ 有资金 $x$,在任意时刻 $s>t$,只要$(b-r)/\sigma^2(1-\alpha)<1$,如果用资金的$\dfrac{b-r}{\sigma^2(1-\alpha)}$部分购买有风险证券,而以其余部分购买无风险证券,那么在用价值函数$\varphi(x)=x^{\alpha}(0<\alpha<1)$作资金评估条件下,在$t_1$时刻平均获益为 $e^{\lambda(t_1-t)}x^{\alpha}$,它是最大获益.

## 6. 期权定价、套期保值与倒向随机微分方程

与控制概念密切相关的随机的终点控制:假定 $b(t,x,v)$充分

好，$\sigma(t,x,v)=v$ 问对将来时刻 $T$ 的一个随机变量 $X$，是否存在 $(\xi_t,\eta_t)$ 及 $x>0$，使 $\xi_0=x$，$\xi_T=X$，$\mathrm{d}\xi_t=b(t,\xi_t,\eta_t)\mathrm{d}t+\eta_t\mathrm{d}B_t$？

这种问题称为倒向随机微分方程. 在数理金融中 $X$ 称为 $T$ 时刻将获得的未定权益（或债权 contingent claim），$\xi_t$ 称为 $X$ 在时刻 $t(<T)$ 的资产价值. [Peng]中证明了对于 $EX^2<\infty$ 情形，上述倒向方程的解 $(\xi_t,\eta_t)$ 唯一存在，于是 $\xi_0(=x$，不再随机）就是在 $t=0$ 时刻购买（$T$ 时获益的）$X$ 所需付的钱（定价）.

例如有一种股票，用 Black-Scholes 模型描述它在时刻 $t$ 的价格 $S_t$：

$$\mathrm{d}S_t = S_t(b\mathrm{d}t + \sigma\mathrm{d}B_t) \quad (b > r > 0),$$

$r$ 为当前的银行利率. 设以这种股票为资产上市一种欧式看涨期权（european call option），即在 $t=0$ 时卖方与买方订了一个合约，规定买方有在时刻 $T$ 以价格为 $K$ 买进一批这种股票的权利（如果当时的市场价格 $S_T$ 低于 $K$，买方可以不买），因此买方在 $T$ 时净得 $X\triangleq(S_T-K)^+$. 由于这个合约能带来 $X$ 的获益，当然需要在 $t=0$ 时用钱来买，问这个合约应该值多少钱才合理呢？设 $\xi_t$ 是 $X$ 在 $t$ 时刻所具有的资产，金融上有一个对冲风险的原则，叫**套期**（hedge）原则，就是在 $t$ 时"虚拟"卖出 $n_t$ 份股票（$n_t$ 可以负，此时表示买进），使余下的部分不再有风险. 也就是"卖出"了 $n_t$ 份后，在 $t+\mathrm{d}t$ 时的资产为 $\xi_{t+\mathrm{d}t}-n_tS_{t+\mathrm{d}t}$，它应该相当于 $t$ 时刻剩余资产 $\xi_t-n_tS_t$ 存进利率为 $r$ 的银行中在 $t+\mathrm{d}t$ 时刻应得的资产，即

$$\xi_{t+\mathrm{d}t} - n_tS_{t+\mathrm{d}t} = (1 + r\mathrm{d}t)(\xi_t - n_tS_t).$$

我们用 $\mathrm{d}\xi_t$，$\mathrm{d}S_t$ 分别代替 $\xi_{t+\mathrm{d}t}-\xi_t$，$S_{t+\mathrm{d}t}-S_t$，上式就变为

$$\mathrm{d}\xi_t - n_t\mathrm{d}S_t = r(\xi_t - n_tS_t)\mathrm{d}t.$$

再用 $\mathrm{d}S_t=S_t(\sigma\mathrm{d}B_t+b\mathrm{d}t)$ 代入，最后得到下述的倒向随机微分方程（下面记 $\eta_t=n_tS_t$，其中 $n_t$ 是未知的）

$$d\xi_t = (\gamma\xi_t + (b - r)\eta_t)\mathrm{d}t + \sigma\eta_t\mathrm{d}B_t,$$

$$\xi_T = X.$$

由于对于满足 $EX^2<\infty$ 的 $X$，方程存在唯一解 $(\xi_t,\eta_t)$，便可自然

地认为$\xi_0$是获得期权$X$这个合约合理的价格. 具体地如何求这倒向随机微分方程的解是一个很为复杂的问题,本节不可能涉及.

**7. 期权定价的 Black-Scholes 公式**

本段介绍 Black-Scholes 在 1973 年的著名论文中关于股票欧式期权定价的主要思想. 设一种与股票相系的期权持有人(乙方)在时刻$T$的收益为$f(S_T)$,股票价格遵从随机微分方程:

$$dS_t = S_t(b dt + \sigma dB_t).$$

问购买这期权在$t=0$时应付多少? 当$f(S_T)=(S_T-K)^+$时这个期权叫欧式看涨(call)期权,当$f(S_T)=(K-S_T)^+$时这个期权叫欧式看跌(put)期权,看跌期权的卖出方有在$T$时刻以$K$的价格从乙方(期权买进方)买进该期权所系股票的义务,但是买进期权方(乙方)可以卖出股票也可以不卖(只在股价跌的时候卖,从中获益$(K-S_T)^+$. 本段与第 6 段是分别从不同的角度去解决同一个问题.

设在$T$收益为$f(S_T)$的股票期权在$t(<T)$时刻价格为$F(t,S_t)$(于是在$t=0$时该期权的定价为$x=F(0,S_0)$)下面要用 Ito 公式求$F(t,x)$的显式表达.

虚拟期权卖出方在$t$时刻买进$\Delta$份($\Delta$待定)股票以抵销在$T$时刻损失$f(S_T)$的风险,即它花费了$\Delta\cdot S_t$,于是在$t$时刻他有余额

$$N_t = F(t,S_t) - \Delta\cdot S_t.$$

到了$t+dt$时刻,这价值变为

$$N_{t+dt} = F(t+dt,S_{t+dt}) - \Delta\cdot S_{t+dt}.$$

于是由 Ito 公式

$$\begin{aligned} dN_t &= dF(t,S_t) - \Delta\cdot dS_t \\ &= \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial x}S_t(b dt + \sigma dB_t) + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}\sigma^2 dt \\ &\quad - \Delta\cdot S_t(b dt + \sigma dB_t) \end{aligned}$$

$$= \left(\frac{\partial F}{\partial x} - \Delta\right)\sigma S_t \mathrm{d}B_t + \left(\frac{\partial F}{\partial t} + bS_t\left(\frac{\partial F}{\partial x} - \Delta\right) + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 F}{\partial x^2}\right)\mathrm{d}t.$$

取 $\Delta = \frac{\partial F}{\partial x}(t, S_t)$ 使得

$$\mathrm{d}N_t = \left(\frac{\partial F}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 F}{\partial x^2}\right)\mathrm{d}t.$$

此时影响随机波动的因素不再出现，所以 $N_t$ 应是无风险的，即

$$\frac{\mathrm{d}N_t}{\mathrm{d}t} = rN_t \quad (r \text{ 为利率}).$$

也就是

$$\frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 F}{\partial x^2} + \frac{\partial F}{\partial t} = r(F - \Delta \cdot S_t) = r\left(F - \frac{\partial F}{\partial x} S_t\right),$$

即 $F(t,x)$ 满足下述 **Black-Scholes 方程**

$$\frac{\partial F}{\partial t} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 F}{\partial x^2} + rx\frac{\partial F}{\partial x} - rF = 0.$$

又由 $F(T, S_T) = f(S_T)$，可知应有终端条件：$F(T,x) = f(x)$. 令 $t' = T - t, x' = \log x, G = \mathrm{e}^{\alpha x' + \beta t'} F$（适当选取常数 $\alpha, \beta$）就变成热方程

$$\frac{\partial G}{\partial t'} = \frac{\partial^2 G}{\partial x'^2}$$

的初值问题，从而可得显式解（Black-Scholes 公式）

$$F(t,x) = \mathrm{e}^{-r(T-t)} \int f(x)\mathrm{e}^{\left(r - \frac{\sigma^2}{2}\right)(T-t) + \sigma y\sqrt{T-t}} \times \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{y^2}{2}} \mathrm{d}y.$$

## 习　题

1. 求对应于相关函数为 $B(s) = \gamma \mathrm{e}^{-\beta|s|}$ $(\gamma, \beta > 0)$ 的非退化的宽马氏平稳 Gauss 过程的条件数学期望 $\mu \triangleq E(\xi_{t+s} | \xi_t = x)$ 与条

件方差

$$E((\xi_{t+s}-\mu)^2|\xi_t=x).$$

2. 上题中若 $\gamma=\beta-1, t_1<t_2<\cdots<t_{m+1}$，记 $\underline{a}_m=(e^{-(t_{m+1}-t_m)},\cdots,e^{-(t_{m+1}-t_1)})$. 证明

(1) $\underline{a}_m=e^{-(t_{m+1}-t_m)}(\underline{a}_{m-1},1)$；

(2) $\underline{a}_m\Sigma_m^{-1}=e^{-(t_{m+1}-t_m)}(1,0,\cdots,0)$； 其中 $\Sigma_m=(e^{-|t_i-t_j|})_{i,j\leqslant m}$；

(3) $\Sigma_{m+1}=\begin{pmatrix}\Sigma_m & \underline{a}_m^T\\ \underline{a}_m & 1\end{pmatrix}$；

(4) $\underline{a}_m\Sigma_m^{-1}\underline{a}_m^T<1$，且

$$\Sigma_{m+1}^{-1}=\begin{pmatrix}\Sigma_m^{-1} & \\ & 0\end{pmatrix}+(1-\underline{a}_m\Sigma_m^{-1}\underline{a}_m^T)^{-1}\begin{pmatrix}(\Sigma_m^{-1}\underline{a}_m^T)\underline{a}_m\Sigma_m^{-1} & -\Sigma_m^{-1}\underline{a}_m^T\\ -\underline{a}_m\Sigma_m^{-1} & 1\end{pmatrix}$$ (用归纳法).

3. 证明 Brown 运动为时齐 Markov 过程，求出它的转移函数.

4. 证明平稳 Gauss-Markov 过程是可逆的，即对任意 $n$ 及 $T>t_n>\cdots>t_1>0$，$(\xi_{t_1},\cdots,\xi_{t_n})$ 总与 $(\xi_{T-t_1},\cdots,\xi_{T-t_n})$ 有相同的联合密度.

5. 仿照定义 9.2 给出取值于 $R^d$ 的连续时间时齐 Markov 过程的定义(假定有联合分布密度).

6. 试证明 $d$ 维的 Brown 运动 $\{\vec{B}_t;t\geqslant 0\}$,

$$\vec{B}_t=\begin{pmatrix}B_t^{(1)}\\ \vdots\\ B_t^{(d)}\end{pmatrix}$$

(参见 §2.3 的第一段)是时齐的 $d$ 维 Markov 过程(即取值为 d

维向量值的 Markov 过程).

7. 设矩阵 $A$ 为非负定对称(即所有特征值都小于 0),$\{\vec{\xi}_t, t\in R\}$ 为 $d$ 维平稳 Gauss 过程,且有

$$E\,\vec{\xi}_t \equiv 0,\quad E(\vec{\xi}_t\,\vec{\xi}_s^T) = \mathrm{e}^{A|s-t|}$$

$\left(\text{矩阵函数 } \mathrm{e}^X \triangleq I + \dfrac{X}{1!} + \dfrac{X^2}{2!} + \cdots\right)$.

(1) 求证 $\{\vec{\xi}_t, t\in R\}$ 的转移密度函数为

$p(t,\vec{x},\vec{y}) = N(\mathrm{e}^{At}\vec{x}, I - \mathrm{e}^{2At})$(其中 $I$ 为单位矩阵);

(2) $\lim\limits_{t\to\infty} p(t,\vec{x},\vec{y}) \to N(0,I)$;

(3) $\{\vec{\xi}_t, t\in R\}$ 是时齐的 Markov 过程.

8. 设

$$\begin{cases} \mathrm{d}\zeta_t = 0 & (\text{一维状态方程}),\\ \mathrm{d}\eta_t = \zeta_t\mathrm{d}t + \gamma\mathrm{d}B_t & (\text{一维测量方程}). \end{cases}$$

求证:若取 $V_0 = \sigma_0^2$, $\hat{\zeta}_0 = 0$,则 $V_t = \dfrac{\sigma_0^2\gamma^2}{\gamma^2 + \sigma_0^2 t}$, $\hat{\zeta}_t = \dfrac{\sigma_0^2}{\gamma^2 + \sigma_0^2 t}\eta_t$.

9. 一维滤波系统

$$\begin{cases} \mathrm{d}\zeta_t = b\mathrm{d}B_t^{(1)},\\ \mathrm{d}\eta_t = \zeta_t\mathrm{d}t + \gamma\mathrm{d}B_t^{(2)} & (B_t^{(1)},\ B_t^{(2)}\ \text{独立}), \end{cases}$$

$V_0 = \sigma_0^2$, $\hat{\zeta}_0 = 0$,证明 $V_t \to \gamma b\ (t\to\infty)$. 又若取 $V_t = 0$, $b = \gamma = 1$,则有 $V_t = \mathrm{th}\,t$, $\hat{\zeta}_t = \dfrac{1}{\mathrm{ch}\,t}\displaystyle\int_0^t \mathrm{sh}\,u\,\mathrm{d}\eta_u$.

10. 一维系统

$$\begin{cases} \mathrm{d}\zeta_t = a\zeta_t\mathrm{d}t,\\ \mathrm{d}\eta_t = \zeta_t\mathrm{d}t + \gamma\mathrm{d}B_t, \end{cases}$$

$E\zeta_0 = c$, $\mathrm{Var}\zeta_0 = \sigma_0^2$,若取 $V_0 = \mathrm{Var}\zeta_0$,求证 $V_t \to 2a\gamma^2$. 再若取 $V_0 = 2a\gamma^2$,求 $\hat{\zeta}_t$.

11. 用 Kalman Bucy 滤波器,把未知参数 $\theta$ 作为状态,对以下

系统求 $\theta$ 的 Kalman-Bucy 滤波

$$d\eta_t = \theta M_t dt + \Gamma_t dB_t \quad \text{（量测方程）}.$$

12. 对一维系统

$$\begin{cases} d\zeta_t = a\zeta_t dt + b dB_t^{(1)}, \\ d\eta_t = m\zeta_t dt + \gamma dB_t^{(2)} \end{cases} \quad (B_t^{(1)}、B_t^{(2)} \text{ 独立}),$$

取 $V_0=\sigma_0^2$，求证 $V_t=\dfrac{\alpha-k\beta\theta_t}{1-k\theta_t}$，其中 $k=\dfrac{\sigma_0^2-\alpha}{\sigma_0^2-\beta}$，

$$\alpha = m^{-2}(a\gamma^2 - \gamma\sqrt{a^2\gamma^2+m^2b^2}),$$

$$\beta = m^{2}(a\gamma^2 + \gamma\sqrt{a^2\gamma^2+m^2b^2}),$$

$$\theta_t = \exp\left\{\frac{\beta-\alpha}{\gamma^2}m^2 t\right\}.$$

又若 $\beta>\alpha$，并取 $V_0=\beta$，则有

$$\hat{\zeta}_t = \hat{\zeta}_t e^{-\delta t} + \frac{m\beta}{\gamma^2}e^{-\delta t}\int_0^t e^{\delta s} d\eta_s,$$

其中 $\delta = \dfrac{\sqrt{a^2\gamma^2+m^2b^2}}{\gamma}$.

# 索　引

## A



## B



## C



## D

## E



## F



## G



## H



## I

**J**



**K**



**L**



**M**



**N**



**O**



**P**

**Q**



**R**



**S**



**T**



**V**

**W**



**X**



**Y**



**Z**

# 参考文献

[Ar] L. Arnold, Stochastic Differential Equations, Theory and Application, John Wiley & Sons, 1974.

[B] P. Billinsley, Ergodic Theory and Information, John Wiley & Sons, 1965.

[BhW] R. M. Bhattacharya and E. C. Waimire, Stochastic Processes with Applications, John Wiley & Sons, 1990.

[Chung] 钟开莱,《初等概率论附随机过程》,人民教育出版社,1980.

[F] W. Feller, An Introdution to Probability Theory and its Application, Vol. I,1957, Vol. II,1971, John Wiley & Sons.

[Fu] Wayne A. Fuller, Introduction to Statistical Time Series, Second Edition, John Wiley & Sons, 1996.

[G] 龚光鲁,《随机微分方程引论》,北京大学出版社,1995.

[Ge] H. O. Georgii, Gibbs Measures and Phase Transitions, Walter de Guyter, 1988.

[GW] G. Grimmett and D. Welsh, Probability an introduction, clarendon Press, Oxford, 1986.

[He] 何声武,《随机过程论》,华东师范大学出版社,1989.

[IM] D. Isaacson and R. W. Madsen, Markov Chains, Theory and Applications, John Wiley & Sons, 1976.

[KS] I. Karatzas and S. E. Shreve, Brownian Motion and Stochastic Calculus, Springer-Verlag, 1990.

[KT] S. Karlin and H. M. Taylor, A First (Second) Course in Stochastic Processes, Academic Press, 1975(1981).

[KSM] J. G. Kemeny and J. L. Snell, Finite Markov Chains, 1983.

[L] J. Lamperti, Stochastic Processes, Springer-Verlag, 1977.

[Ok] B. Oksendal, Stochastic Differential Equations, An Introduction with Applications, Springer-Verlag, 1989.

[Peng] 彭实戈,倒向随机微分方程及其应用,《数学进展》,第 26 卷第 2 期

99-122 页，1997.

[QG] 钱敏平，龚光鲁，《随机过程论》，北京大学出版社，1997.

[R] S. Ross, Stochastic Processes, John Wiley & Sons, 1983.

[RY] D. Revuz and M. Yor, Continuous Martingales and Brownian Motion.

[Sch] Z. Schuss, Theory and applications of Stochastic Differential Equations, John Wiley & Sons, 1980.

[T] H. C. Tijms, Stochastic Models, An Algorithmic Approach. John Wiley & Sons, 1994.

[W] 王梓坤，《随机过程论》，科学出版社，1965.

[Wang] 汪嘉冈，《现代概率论基础》，复旦大学出版社，1988.

[Wr] 汪仁官，《概率论引论》，北京大学出版社，1994.

[Wong] E. Wong, Introduction to Random Processes, Springet-Verlag, 1983.

[X] 谢衷洁，《时间序列分析》，北京大学出版社，1990.

[YWL] 严士健，王隽骧，刘秀芳，《概率论基础》，科学出版社，1982.